AD_____

GRANT NUMBER DAMD17-93-J-3044

TITLE: Detection of Germlike and Somatic Mutations in the Neurofibromatosis 1 Gene

PRINCIPAL INVESTIGATOR: Richard M. Cawthon, M.D., Ph.D.

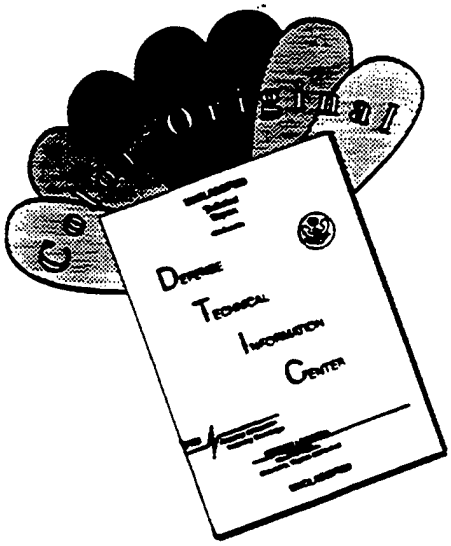
CONTRACTING ORGANIZATION: University of Utah
Salt Lake City, Utah 84112

REPORT DATE: October 1996

TYPE OF REPORT: Final

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF COLOR PAGES WHICH DO NOT REPRODUCE LEGIBLY ON BLACK AND WHITE MICROFICHE.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1996	Final (22 Sep 93 - 21 Sep 96)

4. TITLE AND SUBTITLE Detection of Germline and Somatic Mutations in the Neurofibromatosis 1 Gene

5. FUNDING NUMBERS
DAMD17-93-J-3044

6. AUTHOR(S)
Richard M. Cawthon, M.D., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Utah
Salt Lake City, Utah 84112

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick
Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

This research project sought to improve mutation detection in the neurofibromatosis 1 gene (*NF1*), to determine the spectrum and significance of germline *NF1* mutations in neurofibromatosis 1 patients, and to determine the spectrum and significance of somatic *NF1* mutations in sporadic tumors of several types in the general population. Several mutation detection approaches were explored. The best results came from screening the entire Open Reading Frame of the *NF1* messenger RNA by two rapid, sensitive assays for changes in length, and one additional assay for mutations that prematurely terminate translation of the neurofibromin protein; the combination of these methods resulted in detection of mutations in 19 of 23 unrelated NF1 patients (83%). Since these assays systematically miss mutations in mRNA that result only in amino acid substitutions, it is likely that over 90% of neurofibromatosis 1 - causing mutations alter the structure of the *NF1* mRNA. No differences in the nature or location of *NF1* mutations were discerned when familial and "new mutation" patients with neurofibromatosis 1 were compared. Somatic *NF1* mutations were sought in tumor cell lines established from cancers of the colon, breast, and brain. One or more cell lines representing each tumor type were positive for mutation.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Neurofibromatosis, Mutations	51
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

RC Where copyrighted material is quoted, permission has been obtained to use such material.

RC Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

RC Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

RC For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Richard Carlton 10-21-96

PI - Signature Date

FINAL REPORT FOR GRANT DAMD17-93-J-3044

1. Introduction

1.1 Improving mutation detection in the *NF1* gene (Technical Objective 1)

Mutation detection in *NF1* is difficult due to its large size, approximately 350 kb of genomic DNA, and 11 kb of mRNA. During this period of support from DAMD17-93-J-3044, my laboratory contributed to the final elucidation of all intron-exon boundaries in the *NF1* gene [1] (see attached publication in the Appendix). Also under support from this grant, my laboratory helped identify multiple homologous loci of *NF1* that complicate mutation detection using genomic DNA as template, and we designed primers and reaction conditions for PCR that allow *NF1*-locus specific amplification of every *NF1* exon in genomic DNA despite the presence of these homologous loci [1, 2] (see attached publications in the Appendix). The work of co-author Heidi Huntsman Breidenbach on both these projects was carried out while she was employed as a full-time technician in my laboratory, paid entirely from DAMD17-93-J-3044.

Although these advances in our knowledge of the *NF1* locus at the genomic level have laid the foundation for genomic DNA-based mutation detection in *NF1*, we nevertheless chose to focus our mutation detection efforts on *NF1* mRNA, due to the difficulties of establishing a rapid genomic DNA-based screen of the 60 *NF1* exons, and the pitfalls that allow major classes of mutation (such as intron to intron inversions) to be missed in exon by exon screening at the genomic DNA level [3]. We had shown prior to the beginning of the DAMD17-93-J-3044 funding period that the frequency with which heterozygosity can be detected at polymorphic bp 2034 in the ORF of *NF1* is the same in NF1 patient mRNA, NF1 patient genomic DNA, and control genomic DNA (testing approximately 150 NF1 patients and 100 controls). Therefore, nearly all NF1 patients make detectable levels of mRNA from the disease-causing allele. This finding justified a major effort to increase the speed and sensitivity of detection of mutations in *NF1* mRNA.

1.1.1 *Pros and cons of the Protein Truncation Test in NF1 mutation detection*

During this DAMD17-93-J-3044 funding period my laboratory developed the first protein truncation test (PTT) for *NF1*, an mRNA-based assay. The general method for the PTT was first described by Sarkar et al. [4]; our approach for the *NF1* gene incorporated some of the modifications of PTT suggested for detecting mutations in the *apc* gene [5]. We further modified the method to take advantage of the newly available long PCR method [6]. Our laboratory was the first to publish data from applying PTT to mutation detection in *NF1* [7]. Those results were described in detail in my 21 January 1995 interim Progress Report for this DAMD17-93-J-3044 grant. We detected mutations in 12 out of 25 unrelated NF1 patients screened. Since then, Dr. Heim's laboratory reported mutation detection in 14 of 21 unrelated NF1 patients screened [8]. Roche Molecular Diagnostics now offers a clinical diagnostic assay based on Dr. Heim's screening method that claims a 65% chance of detecting any given NF1 patient's mutation.

In our recent report of long RT-PCR of the entire *NF1* ORF [9] (see attached publication in the Appendix), the first report of such an achievement, we discuss the limitations of the PTT. Because the RT-PCR products that can be effectively analyzed are limited in size to approximately 2 kb or smaller, the 8.5 kb *NF1* ORF must be amplified as a series of five or six overlapping products. Therefore, medium to large deletions that span the region of overlap of two adjacent target segments for RT-PCR amplification will fail to amplify. In these situations, no product will be made from the RNA of the disease-causing allele; however, product of normal structure will still be made from the mRNA made by the remaining normal allele, and so the results of the assay will appear entirely normal. For a great

variety of medium-sized to large deletions, the long RT-PCR assay of *NF1* mRNA [9] will be effective in mutation detection, whereas the PTT will not. During her work on this project Jennifer Martinez was a full-time technician paid entirely from DAMD17-93-J-3044.

The PTT will also fail to detect small, in frame deletions and insertions because the mutant proteins that are synthesized will be negligibly altered in length and so fail to resolve from the normal length proteins made from the normal *NF1* mRNA. PTT will also fail to detect missense mutations, which alter amino acid sequence without altering protein length.

Furthermore, despite various modifications, the PTT is plagued with the problem that multiple polypeptides of various lengths, all shorter than the protein predicted by the normal ORF, may be synthesized depending on the particular segment being screened, even when the template is prepared from a homozygous control. These "extra bands," most likely due to initiation of translation at multiple in-frame methionine codons downstream of the most 5' methionine codon (the codon purposely designed to be the initiator), lower the signal-to-noise ratio in the gel lanes. Many faint mutant bands are likely to be missed due to overlap with normal extra bands. A related problem with the PTT is that when an abnormality in a patient sample is detected, it often consists of more than one mutant band in the gel lane. Presumably, the length of the mutant polypeptide with the highest molecular weight corresponds to the distance from the initiator methionine codon (at the 5' end of the PCR product) to the stop codon that signals a premature stop, and the shorter mutant polypeptides are due to intiation of translation at downstream in-frame methionine codons with continued translation until the same mutant premature stop is encountered. Therefore, to calculate the approximate location of the truncating mutation in the 2 kb segment being screened, one must find the mutant band of highest molecular weight. Since this band is often obscured by normal "extra bands," with this strategy one will often choose the wrong 300-500 bp section of the 2 kb segment to sequence in hunting for the mutation at the sequence level.

1.1.2 *Comprehensive screening for length abnormalities in the NF1 ORF in RT-PCR products*
Fully 63% of the NF1 mutations reported to the NF1 Genetic Analysis Consortium (October 31, 1995) should result in deletions or insertions of various sizes in the *NF1* mRNA (115 mutations out of 182 total mutations reported). Since essentially any length abnormality should be easily detected by conventional agarose gel and polyacrylamide gel electrophoresis, with no "extra bands" expected from control samples, and the expected overall sensitivity for detecting NF1 mutations by this approach (63%) is expected to be similar to the overall sensitivity of the PTT approach (50-67%), we decided to explore a screening algorithm that would begin simply with the testing of *NF1* mRNA for length abnormalities. Let us refer to this approach as Length of ORF (LOO) screening. This algorithm would look first for medium to large deletions and insertions, then for smaller deletions and insertions, down to a single missing or extra base. Only those NF1 patient samples that pass these assays for length abnormalities would then be subjected to assays sensitive to single base substitutions such as detection of Single Strand Conformation Polymorphisms (SSCP) [10], Restriction Endonuclease Fingerprinting (REF) [11], or the PTT.

1.2 Spectrum and significance of germline *NF1* mutations (Technical Objective 2)
This project addressed the following questions regarding the *NF1* mutations that cause neurofibromatosis 1 disease. Do the germline *NF1* mutations of patients with hereditary neurofibromatosis 1 differ in nature or location from the constitutional *NF1* mutations of sporadic (or "new mutation") NF1 patients? What fraction of unrelated NF1 patients can be shown to have an abnormality of *NF1* mRNA structure?

1.3 Spectrum and significance of somatically acquired *NF1* mutations (Technical Objective 3)
This project addressed the following questions regarding the *NF1* mutations of somatic cells that contribute to the development of sporadic cancers in the general population (individuals without neurofibromatosis 1 disease). In which types of sporadic tumors does somatic mutation of *NF1* play a role? Is the spectrum of somatic mutations in *NF1* in sporadic tumors different from the spectrum of germline mutations that cause neurofibromatosis 1 disease? Is the spectrum of somatic mutations in *NF1* different in different tumor types, suggesting different roles for *NF1* in different tissues?

We revised the original strategy proposed in DAMD17-93-J-3044 to investigate the role of somatic *NF1* mutations in the development of sporadic tumors. Many tumor cell lines derived from cancers of various types are available from cell and tissue banks. RNA can be readily prepared from these cell lines, with no contamination from normal tissue, and the entire *NF1* ORF rapidly screened. Since somatic mutations in cell lines may arise during culture, the finding of an *NF1* mutation in a tumor cell line should not be considered definitive evidence that a somatic *NF1* mutation contributed to that tumor's development *in vivo*. Nevertheless, it seems likely that the relative frequencies of *NF1* mutations in different tumor types *in vivo* will be roughly reflected in cell lines established from tumors of those types. Thus, the data collected from surveying tumor cell lines for *NF1* mutations should allow us to prioritize our subsequent searches for *NF1* mutations in never-cultured tumor specimens. Furthermore, if *NF1* mutations found in cell lines representing a particular tumor type turned out to be restricted to a particular segment of the gene, subsequent efforts to screen tumor specimens of that type could reasonably be focused to that segment. Finally, totally negative results obtained from screening cell lines from a particular tumor type would justify omitting the corresponding tumor specimens from subsequent screens.

2. Materials and Methods

2.1 Cell lines
Epstein Barr Virus - transformed lymphoblast lines from NF1 patients were provided by Dr. Ray White and Dr. Ken Ward of the University of Utah. Lymphoblast cell lines were grown in RPMI-1640 medium with 15% fetal bovine serum (FBS) in a humidified 37°C incubator in the presence of 5% CO_2. Tumor cell lines from ATCC (see Table 3) were grown as recommended by ATCC. Glioblastoma cell lines JBSA, PFAT(1), and PFGC(3) were grown in Dr. Fults' laboratory in DMEM + 10% FBS in the presence of 10% CO_2. Colon cancer cell lines were grown in Dr. Markowitz's laboratory, as described in ref. 12 and its citations.

2.2 RNA preparation and first strand cDNA synthesis
Total RNA preparation from all cell lines and whole blood, and first stand cDNA synthesis from total RNA were as described in reference 9 (included in the Appendix), except that RNA preparation from the colon cancer cell lines was performed in Dr. Markowitz's laboratory as described in ref. 12.

2.3 Long RT-PCR of the entire *NF1* Open Reading Frame (ORF)
This method is presented in reference 9 (see Appendix).

2.4 Restriction digestion of long *NF1* RT-PCR product and analysis by agarose gel electrophoresis
This method is presented in reference 9 (see Appendix).

2.5 PCR of *NF1* ORF as five products suitable for mutation screening by multiple methods
Five overlapping, approximately 2 kb long PCR products were synthesized, using the 8.7 kb *NF1* RT-PCR product as template. These 2 kb PCR products are called TnT products because they are also

suitable for Transcription and Translation studies. Table 1 shows the sequences of these primers and the sizes of the products.

For each of the five products, the template for PCR was one µl of a 1:10 dilution in water of the 8.7 kb long *NF1* RT-PCR product. The final components of the PCR were 0.2 mM dNTPs, 1x XL PCR buffer (Perkin-Elmer), 1.1 mM $Mg(OAc)_2$, 0.25 mM spermidine, 0.4 µM of each primer, and 1 unit of Taq DNA polymerase in a final volume of 20 µl. A hot start procedure was followed, using a Perkin-Elmer 9600 thermal cycler: 94°C x 1.5'; 80°C hold while enzyme is added; (94°C x 30", 54°C x 30", 65°C x 4') x 12 cycles; 72°C x 10'.

After PCR, the entire volume of each TnT product was loaded into a single well of a 1-2% agarose gel in a Tris-acetate-EDTA buffer system with ethidium bromide in both the gel and the buffer, and electrophoresis was carried out until the product had migrated 1-2 cm into the gel. The bands containing the TnT products were then visualized over a long wave UV light box and excised with a razor blade. The DNA was extracted from the gel slices by the GeneClean II method (Bio101), according to the manufacturer's instructions. Final elutions were into 10 µl of water.

To radiolabel one strand of each TnT product, the purified products were subjected to one round of denaturation, annealing, and extension in the presence of ^{33}P-dATP and one of the primers from the original 2 kb PCR reaction. For TnT1, TnT2, TnT4, and TnT5 the downstream primer was used for this reaction; for TnT3 the upstream primer was used. The components in the final reaction mixture were 2.5 µl of GeneClean II-purified DNA, 1x XL PCR buffer (Perkin-Elmer), 0.02 mM dNTPs (10 fold lower than for a normal PCR), 1.1 mM $Mg(OAc)_2$, 0.25 mM spermidine, 0.5 units of Taq DNA polymerase, 0.4 µM primer, and 0.8 µl ^{33}P-dATP (1000-3000Ci/mmol, 10mCi/ml, Andotek Life Sciences), in a final volume of 10µl. The sample was denatured at 94°C x 30", annealed at 54°C x 30", and extended at 65°C x 4' (one cycle only).

2.6 Restriction digestion of the five products and analysis by polyacrylamide gel electrophoresis

All restriction endonuclease digestions were performed by sequential additions of reagents to the 10 µl radiolabeled samples synthesized in the step immediately preceding this section. All digestions were performed in the presence of 4mM spermidine. Reaction components and steps specific to each digestion are given below.

2.6.1 *Restriction enzyme digestion of TnT1*

Add 5 units of Mam I (Boehringer-Mannheim, or BM), 5 units of Sfu I (BM), and 1.5 µl of 10x Sfu I buffer (BM) to a final volume of 15 µl; incubate at 37°C x 1 hour. Then add 5 units of Bsr I in a 1.0 µl volume (New England Biolabs, or NEB) and incubate at 65°C for 1 hour.

2.6.2 *Restriction enzyme digestion of TnT2*

Add 5 units of Msp I (GIBCO-BRL) and 1.5 µl of 10x Msp I buffer to a final volume of 15 µl; incubate at 37°C x 1 hour. Then add 5 units of Ava II (GIBCO-BRL), and NaCl to a final concentration of 50 mM, in a final volume of 16.4 µl; incubate 37°C x 1 hour.

2.6.3 *Restriction enzyme digestion of TnT3*

Add 5 units Hae III (GIBCO-BRL) and 1.5 µl of 10x Hae III buffer to a final volume of 15 µl; incubate at 37°C x 1 hour. Then add 5 units of EcoRI (GIBCO-BRL), and NaCl to a final concentration of 50 mM, in a final volume of 16 µl; incubate 37°C x 1 hour.

Table 1

Primer	Sequence	Exons	Name and size (bp) of PCR product	Restriction enzymes	Restriction fragment sizes (bp)
2T7-1	GGATCCTAATACGACTCACTATAGGGAGAggacatggccgcgcac	1	TNT1:		450, 398, 321, 306,
3TNT1	CTACTAATGATGATGATGATGATGccggatatctgcttcctcaca	13	2219	BsrI+MamI+SfuI	277, 266, 201
5TNT2	GGATCCTAATACGACTCACTATAGGGAGACCACCAtggttcttcatcagttagatagca	11	TNT2:		446, 389, 364, 274,
3TNT2	CTACTAATGATGATGATGATGATGAgccaattctacttctttagaaaaca	22	2215	AvaII+MspI	231, 189, 183, 139
5TNT3	GGATCCTAATACGACTCACTATAGGGAGACCACCAtggaattgatggaagccaaatcac	19b	TNT3:		441, 361, 324, 320, 257,
3TNT3	CTACTAATGATGATGATGATGATGgccctggtttgcaatggttaag	29	2140, 2203	HaeIII+EcoRI	225, 200, 159, 149, 24
5TNT4	GGATCCTAATACGACTCACTATAGGGAGACCACCATGgatttgctgatataccatgtctt	28	TNT4:		415, 399, 365, 349,
3TNT4	CTACTAATGATGATGATGATGATGCttccggattgccataaatacttc	39	2296	Cac8I+PstI+TfiI	326, 185, 147, 110
5TNT5	GGATCCTAATACGACTCACTATAGGGAGACCACCATGgctttgacatccttggaaacag	34	TNT5:		449, 391, 359, 275,
3TNT5	CTACTAATGATGATGATGATGATGGaagaaagcaagcaagctAcacac	49	2062	RsaI+AvaI	242, 190, 156

2.6.4 *Restriction enzyme digestion of TnT4*

Add 5 units Cac8 I (NEB), 1.5 µl of 10x Cac8 I buffer, and 6 units of Pst I (NEB) to a final volume of 14.1 µl; incubate at 37°C x 1 hour. Then add 2.5 µl containing 5 units of Tfi I (NEB); incubate at 65°C x 1 hour.

2.6.5 *Restriction enzyme digestion of TnT5*

Add 5 units of Rsa I (GIBCO-BRL) and 1.5 µl of 10x Rsa I buffer to a final volume of 15 µl; incubate at 37°C x 1 hour. Then add 5 units of Ava I (GIBCO-BRL), and NaCl to a final concentration of 50 mM; incubate at 37°C x 1 hour.

2.6.6 *Fill-in of restriction fragment ends*

We found that while the expected radiolabeled single-stranded DNA restriction fragments were observed upon denaturing polyacrylamide gel electrophoresis, each fragment was accompanied by additional fragments one or more bases longer than the expected length. This was prevented by completely removing unincorporated dNTPs following the radiolabeling step, but prior to the addition of restriction enzymes; with this modification, only the shortest (i.e. the expected) version of each restriction fragment was observed. Our interpretation, therefore, is that one or more fragments longer than the one expected from restriction digestion were generated by the addition of dNTPs to the exposed 3' ends of the restriction fragments by the action of Taq polymerase, which remained present and active in the samples. This occurred even for blunt-end restriction fragments, due to Taq's well-known ability to add an extra 3'-overhanging base to double-stranded DNA. However, these base additions did not go to completion, probably because the optimal reaction conditions for restriction digestion were sub-optimal for dNTP incorporation by Taq polymerase. Therefore, mixed length populations of each restriction fragment resulted, which would greatly complicate detection of mutant bands on gels.

As an alternative to completely removing unincorporated dNTPs prior to restriction digestion, we chose to optimize the fill-in and untemplated base addition reactions of Taq polymerase in order to obtain populations of restriction fragments of uniform length (from controls). Following restriction digestion, described above, 400 µl of TE^{-4} (10 mM Tris, 0.1 mM EDTA) was added to each sample. The samples were then loaded onto Microcon-50 spin columns (Amicon, Inc.) and centrifuged at 14,000 x g for 6 minutes. This buffer exchange procedure left the restriction fragments in 5 µl of TE^{-4}. The samples were then brought to a 10 µl final volume containing 0.25 units of Taq polymerase, 0.1 mM dNTPs, 1x XL PCR buffer (Perkin-Elmer), and 1.1 mM $Mg(OAc)_2$, and incubated at 65°C for 10-15 minutes.

2.6.7 *Polyacrylamide gel electrophoresis under denaturing conditions*

Each sample was analyzed on two different polyacrylamide slab gels, one optimized for resolving the larger-sized fragments, the other optimized for resolving smaller fragments; the BRL S2 electrophoresis apparatus was used for all gels, which were cast 38 cm long each. Radiolabeled samples were diluted 1:1 with a solution containing 95% formamide, 20 mM trisodium EDTA (pH approx. 8.8), 0.05% bromphenol blue and 0.05% xylene cyanol. The samples were then heated at 90°C for 3 minutes (or 80°C for 5 minutes) to denature the DNA, and three to five µl were loaded per gel lane. Gel One (for better resolution of the larger fragments) was a 4.75% sequencing gel made with Sequagel-4.75 (National Diagnostics), according to the manufacturer's instructions; electrophoresis was carried out in the presence of an electrolyte gradient[13]. At the beginning of the run, the top buffer chamber contained 0.5xTBE buffer, and the bottom chamber contained 1xTBE. After electrophoresis at 85W constant

power for 2 hours, the power was turned off and the bottom buffer was brought to 1M NaAc by the addition of ½ volume of 3M NaAc; electrophoresis was then continued for 2 additional hours at 85 W. Gel Two (for better resolution of the smaller fragments) was a conventional 6% gel made with Sequagel-6 (National Diagnostics). Both top and bottom buffer chambers contained 1xTBE, and electrophoresis was for 2 hours at 85 W. After electrophoresis, one glass plate was removed, and the gels were transferred to 3MM Whatman blotting paper, covered with plastic wrap, and dried over heat under vacuum on a Hoeffer slab gel dryer. The plastic wrapped was then removed, the dried gel exposed to Kodak X-Omat Scientific Imaging Film at -80°C in the dark for 1-5 days, and the exposed film developed in a Kodak film developer.

3. Results and Discussion

3.1 Improving mutation detection in the *NF1* gene (Technical Objective 1)

3.1.1 *Detection of large scale deletions in NF1 mRNA*

Our recent report on long RT-PCR of the entire *NF1* ORF and analysis of the products on agarose gels, with and without restriction enzyme digestion [9] (see Appendix) demonstrated the utility of this approach in detecting large scale, multi-exon deletions in *NF1* mRNA. In principle this technique should be able to detect even mutant *NF1* mRNAs resulting from genomic DNA deletions > 100 kb in size; the only requirements are that the deletion must lie between the sites where the PCR primers bind (exon 1 and 49), and a stable mRNA must be made.

3.1.2 *Detection of small scale deletions and insertions in NF1 mRNA*

Figure 1 illustrates the strategy for detecting small deletions and insertions in the *NF1* ORF using cDNA as template. First, five overlapping, approximately 2 kb PCR products were synthesized, using the long *NF1* RT-PCR product as template. These 2 kb PCR products (called TnT products because they are also suitable for Transcription and Translation studies) were then subjected briefly to agarose gel electrophoresis, followed by excision and purification from the agarose to remove unincorporated PCR primers. Next, one strand was radiolabeled by performing a single round of denaturation, annealing, and extension with ^{33}P-dATP present, using just one of the two primers present during the original 2 kb PCR. The resulting radiolabeled products were then cut with restriction enzymes, carefully selected for each specific TnT product based on its known restriction map, and subjected to electrophoresis in denaturing polyacrylamide gels. Figure 1 shows scanned autoradiographs of gel lanes containing restriction-digested TnT products from normal segments of the *NF1* ORF.

To test whether single base length differences were being resolved, two sets of size standards differing in length by one base were used. The first set consisted of two PCR products, 449 bp and 448 bp long, identical in sequence except for the presence of an extra base at one end of the longer product. Radiolabeled in one strand only, these products were routinely loaded separately and as a mixture on each 4.75% gel. In Figure 2 these standards are shown after migrating approximately halfway through the gel (i.e. after migrating approx. 19 cm). This scanned autoradiograph has lower resolution than the actual film, which does allow two distinct bands to be seen in the lane containing the mixture.

The second set of size standards for testing the resolution of single base length differences was generated by performing the restriction digestion of the radiolabeled TnT3 product in the absence of dNTPs and, separately, in the presence of dNTPs and Taq DNA polymerase. Figure 3, lane a shows two fragments from TnT3 after digestion in the absence of dNTPs; lane b shows the same fragments after digestion in the presence of dNTPs and Taq polymerase; and lane c shows the mixture of the two sets of digestion products. Both of these restriction cuts (with Hae III) should generate blunt ends;

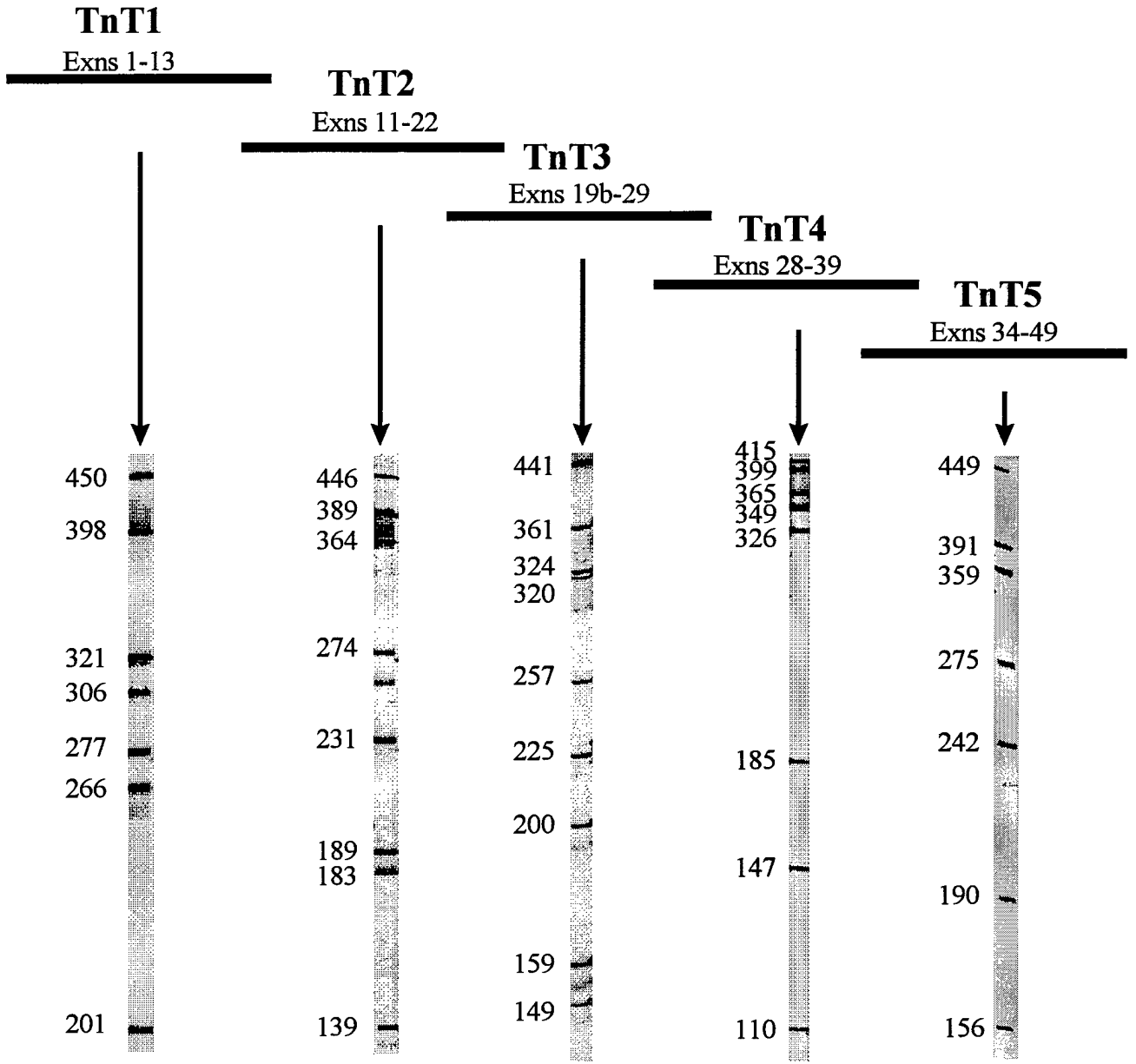

TnT1
Exns 1-13
TnT2
Exns 11-22
TnT3
Exns 19b-29
TnT4
Exns 28-39
TnT5
Exns 34-49
450
398
321
306
277
266
201
446
389
364
274
231
189
183
139
441
361
324
320
257
225
200
159
149
415
399
365
349
326
185
147
110
449
391
359
275
242
190
156

FIGURE 1

449 448 mix

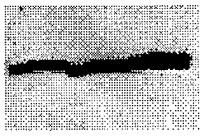


Figure 2

a b c

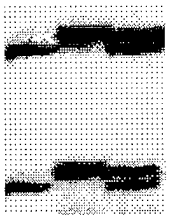


Figure 3

therefore, Taq is not filling in recessed 3' ends here. Rather, in the samples loaded in lane b, it has added a single extra base to the exposed 3' terminus of the radiolabeled strand. This result demonstrates that restriction fragments differing in length by one base are resolved as distinguishable bands in this region of this gel. The 441, 324, 320, 257, 200, 159, and 149 bp restriction fragments from TnT3 are all blunt-ended at the 3' end of the radiolabeled strand, and so are suitable for testing single base length difference resolution by this approach.

3.1.3 *Detection of single base substitutions in NF1 mRNA*

This project did not have sufficient time to perform comprehensive screening for single base substitutions in *NF1* mRNA, but the five TnT RT-PCR products that we have designed are suitable for such studies. First, the radiolabeled restriction fragments from each digested TnT product can be heat-denatured, then subjected to SSCP and/or heteroduplex analyses in non-denaturing polyacrylamide gels with and without glycerol (see protocols from the AT Biochem division of FMC corportation). Alternatively, or in addition, the TnT products can be used for the Protein Truncation Test, since the 5' ends of each TnT product have been designed to allow transcription and subsequent translation of the transcripts [5]. Furthermore, the 3' ends of each TnT product have been designed so that all synthesized polypeptides that reach the end of the normal reading frame should acquire a run of histidine residues that in principle should allow their removal on a nickel-based affinity matrix [14]. The prematurely terminated mutant polypeptides, which should not bind this matrix, can then be eluted and analyzed by SDS-PAGE. This approach should tremendously increase the signal-to-noise ratio for mutation detection by the PTT. (Note that we have introduced a single base change near the 3' end of TnT5 so that protein made from the normal *NF1* sequence will not terminate at the normal *NF1* stop codon, but will read through and still acquire the primer-encoded histidines). However, initial efforts to remove full-length proteins by this approach have been unsuccessful (personal communication, Dr. David Viskochil). An alternative approach, yet to be explored to our knowledge, would use antibodies to some other specific, PCR-primer-encoded carboxy-terminal peptide in an entirely analogous scheme to remove the proteins that are synthesized based on the normal *NF1* mRNA.

3.2 Spectrum and significance of germline *NF1* mutations (Technical Objective 2)

Our data from studies of germline *NF1* mutations are summarized in Table 2. RNA samples from 20 unrelated patients with hereditary neurofibromatosis 1 and 13 unrelated patients with sporadic (or "new mutation") neurofibromatosis 1 were analyzed. By performing the long RT-PCR of *NF1* mRNA only on patients already known (by testing of genomic DNA) to be heterozygous at bp 702 in exon 5, we produced RT-PCR products that could be tested for representation of both alleles. All of the long RT-PCR products prepared from these patients showed heterozygosity at bp 702 by an RsaI restriction enzyme digestion-based assay. The 33 NF1 patients were screened for length abnormalities occurring anywhere in the 8.5 kb *NF1* ORF, by the methods described above. Six of the 33 showed abnormalities in the AseI + AviII restriction digestion products released from the 8.7 kb long RT-PCR product. Each of these, upon sequencing, turned out to be due to a single exon drop-out, except for the mutation of patient 119-1, which was a 30 bp deletion from exon 9. Eleven additional patients showed abnormalities when the five TnT products were digested with segment-specific restriction enzymes and analyzed by PAGE under denaturing conditions. This gives a total detection rate, using length-of-ORF (LOO) screening, of 17/33 (about 52%). This is slightly higher than our detection rate with the Protein Truncation Test alone, 11/23 (about 48%). Looking only at the 23 patients who were screened by both LOO and PTT gives us an overall detection rate of 19/23 (about 83%). We think it is likely that further screening of these *NF1* RNA samples by other methods is likely to identify mutations in some or all of

Table 2

NF1 Patient	familial or sporadic	long RT-PCR product: heterozygosity at bp 702	long RT-PCR product: AGE, uncut	long RT-PCR product: AGE after cutting with AseI + AvrII	5 overlapping TnT PCR products, restriction digested and analyzed by denaturing PAGE: TnT1, cut with BsrI, MamI, SfuI [Exns 1-13]	TnT2, cut with AvaII, MspI [Exns 11-22]	TnT3, cut with HaeIII, EcoRI [Exns 19b-29]	TnT4, cut with Cac8I, PstI, TfiI [Exns 28-39]	TnT5, cut with RsaI, AvaI [Exns 34-49]	abnormal restriction fragment size?	abnormal PTT?	Any abnormality detected?	Mutation description
87-3	F	+	-	-	-	-	-	-	+	+	-	+	
92-5	F	+	-	-	-	-	-	-	-	-	+	+	del 1 bp, 7891, exn 45
94-4	F	+	+	+	-	-	-	+	-	+	+	+	del exn 29, fs, ag->ac ending intrn 28
107-6	F	+	-	+	-	+	+	-	-	+	-	+	del exn 20, fs, ag->ac ending intrn 19
114-3	F	+	-	-	-	-	-	-	-	-	-	-	
117-1	F	+	-	+	+	-	-	-	-	+	not done	+	del exn 3 (no frameshift)
119-1	F	+	-	+	+	-	-	-	-	+	-	+	del 30bp, exn 9, 1230-1259
120-4	F	+	-	-	-	-	-	-	-	-	not done	-	
124-2	F	+	-	-	-	-	?	-	-	+	-	+	
11334	F	+	-	-	-	-	-	-	-	-	+	+	
11345	F	+	-	-	-	+	-	+	-	+	+	+	del 7 bp, 3183-3189, exn 19a
11348	F	+	-	-	-	-	-	-	-	-	not done	-	
11357	F	+	-	-	-	-	-	-	-	-	not done	-	
11388	F	+	-	-	-	-	-	-	-	-	-	-	
11478	F	+	-	-	-	-	-	-	-	-	not done	-	
11547	F	+	-	-	-	-	-	-	+	+	+	+	
11561	F	+	-	-	-	+	+	-	-	+	not done	+	
11562	F	+	-	-	-	-	-	+	-	+	-	+	
11571	F	+	-	-	-	-	-	+	-	+	-	+	
11626	F	+	-	-	+	-	-	-	-	+	not done	+	
11387	S	+	-	+	-	-	-	+	+	+	-	+	del exn 37 (no frameshift)
11390	S	+	-	-	-	-	-	-	-	-	not done	-	
11403	S	+	-	-	+	-	-	-	-	+	not done	+	
11439	S	+	-	-	+	-	-	-	-	+	-	+	del 4bp, 57-60, exn 1
11444	S	+	-	-	-	-	-	-	-	-	not done	-	
11480	S	+	-	-	+	-	-	-	-	+	+	+	del 4bp, 495-498, exn 4b
11553	S	+	-	-	-	-	-	-	-	-	+	+	cga->tga, 1316, exn 10a
11554	S	+	-	-	-	-	-	-	-	-	+	+	tca->tga, 1094, exn 8
11559	S	+	-	-	-	-	-	-	-	-	+	+	truncating mutation around bp 4000
11566-1	S	+	-	-	-	-	-	-	-	-	-	-	
11567	S	+	-	-	-	-	-	-	-	-	+	+	cag->tag, 3210, exn19b
11621	S	+	-	-	-	-	-	-	-	-	-	-	
12359	S	+	-	+	-	-	-	+	+	+	+	+	del exn 35, fs, gt->ct beginning intrn 35
TOTALS		33/33	1/33	6/33	6/33	3/33	3/33	6/33	4/33	17/33	11/23	23/33	14/33

the patients showing no abnormalities so far. For example, single base substitution missense mutations that change an amino acid are known to be a cause of neurofibromatosis 1 disease, but would be systematically missed by both LOO and PTT. Although my laboratory will not be able to perform any further work to confirm and further define at the sequence level the abnormalities reported here, nor be able to search for additional mutations in this patient set by other methods, due to my terminating all my work related to the *NF1* gene, I will give a copy of this report and all NF1 patient samples to Dr. David Viskochil (who continues to work on neurofibromatosis 1) at the University of Utah, so that he may further investigate the mutations in these patients.

Sub-localizing the site of a mutation's underlying sequence change within the affected 2 kb TnT RT-PCR product, based only on an analysis of the mutant band first seen by denaturing PAGE of that TnT product's restriction digest, is frequently not straightforward. In a gel lane if the band corresponding to a particular normal restriction fragment is clearly diminished in intensity relative to other normal bands, then the mutation must affect that restriction fragment in the mutant allele. Thus, one looks for a new band of abnormal mobility, but also for changes in the intensity of bands of normal mobility. Unfortunately, more than 80% of NF1 patients show unequal expression from the two *NF1* alleles [15], and in our experience, in every case where we identified a patient's mutation in RNA and unequal expression was present, the mRNA from the mutant allele was diminished in abundance relative to that from the normal allele. Therefore, in many cases a faint mutant band of abnormal mobility is found in these LOO screens, but no obvious diminution in the intensity of a normal band accompanies it to help narrow the site of mutation, because the % change in signal intensity is small. Nevertheless, the screening of the TnT digestion products achieves the narrowing of the mutation location to one 2 kb segment of the 8.5 kb ORF. Our approach is then to prepare four or five overlapping PCR products (each 400-500 bp long) spanning the 2 kb segment, analyze them on agarose gels (and, if necessary, polyacrylamide gels), identify the new mutant product, and sequence it directly.

The problem of detecting single base deletions and insertions deserves special attention. In Table 2, patient 92-5 is the only patient known to have a single bp deletion, and it was not detected by the LOO approach. We suspect that many single base deletions and insertions may be missed on these gels, especially in the higher molecular weight restriction fragments. As suggested by Figure 2, single base length differences in these relatively large fragments are only barely discernible. Since more peak-to-peak separation is required to distinguish two bands of unequal intensity than to distinguish two bands of equal intensity, sensitive screening of *NF1* for single base deletions and insertions, when unequal expression is the rule, will be difficult to achieve by this approach. Ways around this problem, which we have not had time to investigate, include using other sets of restriction enzymes, in addition to those recommended here, to cut the 2 kb PCR products, so that the segments previously located in the largest restriction fragments will now be in smaller fragments, for which resolution is better. Another solution is to routinely heat and cool the digested products to drive heteroduplex formation, then subject them to PAGE on non-denaturing gels. Heteroduplex-based mutation detection is thought to be nearly 100% sensitive in the detection of single base deletions and insertions.

The germline *NF1* mutations of patients with hereditary neurofibromatosis 1 did not clearly differ in nature or location from the constitutional *NF1* mutations of sporadic (or "new mutation") NF1 patients, in this patient set.

3.3 Spectrum and significance of somatically acquired *NF1* mutations (Technical Objective 3)

Table 3 summarizes the results of screening for somatic mutations in the *NF1* gene in several tumor cell lines established from sporadic cancers of various types arising in individuals without

Table 3

Type of cancer	Cell line	Source	long RT-PCR: AGE, uncut	long RT-PCR: AGE after cutting with AseI + AviII	5 overlapping TnT PCR products, restriction digested and analyzed by denaturing PAGE: TnT1, cut with BsrI, MamI, SfuI [Exns 1-13]	TnT2, cut with AvaII, MspI [Exns 11-22]	TnT3, cut with HaeIII, EcoRI [Exns 19b-29]	TnT4, cut with Cac8I, PstI, TfiI [Exns 28-39]	TnT5, cut with RsaI, AvaI [Exns 34-49]	Mutation description
breast adenocarcinoma	MCF7	ATCC	-	-	?	?	-	-	?	
breast ductal carcinoma	T-47D	ATCC	-	?	?	-	-	-	-	
colon adenocarcinoma	11	Markowitz	-	-	-	-	-	+	-	aag->aaa (Lys->Lys) at bp 5172 (destroys restriction site)
colon adenocarcinoma	12	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	13	Markowitz	-	+	-	-	-	+	+	del exn 37 (no frameshift)
colon adenocarcinoma	981	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	985	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	1025	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	1042	Markowitz	-	-	-	-	?	-	-	
colon adenocarcinoma	1103	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	1211	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	1782	Markowitz	-	+	+	-	-	-	-	del exn 7 (no frameshift)
colon adenocarcinoma	1885	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	2050	Markowitz	-	-	-	-	+	-	-	
colon adenocarcinoma	2082	Markowitz	-	-	-	-	?	?	-	
colon adenocarcinoma	2111	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	2207	Markowitz	-	-	+	-	-	-	-	
colon adenocarcinoma	2213	Markowitz	-	-	-	-	-	-	-	
colon adenocarcinoma	2045a	Markowitz	-	+	+	-	-	-	-	del exn 7 (no frameshift)
glioblastoma multiforme	JBSA	Fults	-	?	-	-	+	-	-	
glioblastoma multiforme	PFAT(1)	Fults	-	-	-	-	+	-	-	del exns 27a, 27b (no frameshift)
glioblastoma multiforme	PFGC(3)	Fults	-	-	?	-	-	-	-	
glioblastoma	U-138 MG	ATCC	-	-	?	-	-	-	-	
glioblastoma, astrocytoma	U-373 MG	ATCC	-	-	-	-	-	-	-	
glioblastoma, astrocytoma	U-87 MG	ATCC	-	+	-	-	+	-	-	
neuroblastoma	IMR-32	ATCC	-	-	-	-	-	-	?	
neuroblastoma	SK-N-SH	ATCC	-	-	-	-	-	-	-	

neurofibromatosis 1 disease. Unfortunately, there was not sufficient time to confirm and further define many of these apparent mutations at the sequence level. In interpreting these results one should keep in mind the possibility that somatic *NF1* mutations may occur in 0, 1, or 2 alleles in any given tumor cell line (unlike the case for constitutional *NF1* mutations in patients with neurofibromatosis 1). For example, breast adenocarcinoma MCF7 appeared to have abnormalities in at least two well-separated segments of the *NF1* mRNA; this is consistent with the possibility that one mutation is present in mRNA from one allele and a different mutation is present in mRNA from the other allele.

Anomalous bands on gels and/or sequence changes suggestive of abnormal *NF1* function were found in 2/2 breast carcinoma cell lines, 7/17 colon cancer cell lines, 5/6 glioblastoma multiforme cell lines, and 1/2 neuroblastoma cell lines. RNA from these cell lines have been screened to date only by the Length of ORF screen. Exon dropouts were confirmed at the cDNA sequence level in three colon cancer cell lines and one glioblastoma multiforme cell line. (It should be noted that the colon cancer cell lines selected for *NF1* screening were specifically chosen to be lines that were negative for activating mutations of the *K-ras* protooncogene.) Although my laboratory will not be able to perform any further work to confirm and further define at the sequence level the abnormalities reported here (due to my terminating all my work related to the *NF1* gene), I will notify Drs. Markowitz and Fults of these results so that they can pursue these findings.

These results suggest that *NF1* undergoes somatic mutation frequently and contributes to a variety of tumor types. Clearly, much more screening is needed to get reliable estimates of *NF1* mutation frequencies in tumor cell lines, and ultimately in fresh tumor specimens. There is currently insufficient data to determine whether the spectrum of somatic mutations in *NF1* in sporadic tumors differs significantly from the spectrum of germline mutations that cause neurofibromatosis 1 disease, and whether the spectrum of somatic mutations in *NF1* varies significantly between tumor types suggesting different roles for *NF1* in different tissues.

4. Conclusions

This research project had three technical objectives. Achieving the second and third objectives with large numbers of samples depends strongly on achieving the first objective. Because of the paramount importance of developing efficient mutation detection in the *NF1* gene, so that other investigators will be able to make rapid progress in several areas of *NF1* research, I chose to focus nearly all my effort on the first technical objective. I believe the Length of ORF assay that I have developed is a major contribution to the field. If this assay is applied first, followed by assays for single base substitution mutations, it is likely that more than 90% of all neurofibromatosis 1 - causing mutations can be identified rapidly. Somatic mutations in *NF1* appear to be common in the tumor types investigated here. Further work by others will be needed to provide an accurate assessment of the frequency and significance of somatic *NF1* mutations in human tumors. One hopes that eventually, an assay that tests the integrity of all neurofibromin functions relevant to neurofibromatosis 1 disease and spontaneous tumorigenesis will be developed that obviates the types of mutation screening that I have developed.

5. References

1. Li, Y. *et al.* Genomic organization of the neurofibromatosis 1 gene (NF1). *Genomics* **25**, 9-18 (1995).
2. Purandare, S.M. *et al.* Identification of Neurofibromatosis 1 (*NF1*) Homologous Loci by Direct Sequencing, Fluorescence *in Situ* Hybridization, and PCR Amplification of Somatic Cell Hybrids. *Genomics* **30**, 476-485 (1995).
3. Lakich, D., Kazazian, H., Jr., Antonarakis, S.E. & Gitschier, J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A [see comments]. *Nat Genet* **5**, 236-41 (1993).
4. Sarkar, G. & Sommer, S.S. Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. *Science* **244**, 331-334 (1989).
5. Powell, S.M. *et al.* Molecular diagnosis of familial adenomatous polyposis. *N Engl J Med* **329**, 1982-7 (1993).
6. Cheng, S., Fockler, C., Barnes, W.M. & Higuchi, R. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91**, 5695-5699 (1994).
7. Cawthon, R. & Breidenbach, H.H. Identification Of *NF1* Mutations By A Protein Truncation Assay. *Am J Hum Genet* **55**, A216 (1994).
8. Heim, R.A. *et al.* Distribution of 13 truncating mutations in the neurofibromatosis 1 gene. *Hum Mol Genet* **4**, 975-81 (1995).
9. Martinez, J.M., Breidenbach, H.H. & Cawthon, R.M. Long RT-PCR of the entire 8.5 Kb *NF1* Open Reading Frame and mutation detection on agarose gels. *Genome Research* **6**, 58-66 (1996).
10. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874-879 (1989).
11. Liu, Q. & Sommer, S.S. Restriction endonuclease fingerprinting (REF): a sensitive method for screening mutations in long, contiguous segments of DNA. *Biotechniques* **18**, 470-7 (1995).
12. Leach, F.S. *et al.* Amplification of cyclin genes in colorectal carcinomas. *Cancer Res* **53**, 1986-9 (1993).
13. Sheen, J.-Y. & Seed, B. Electrolyte Gradient Gels for DNA Sequencing. *BioTechniques* **6**, 942-944 (1988).
14. Kubalek, E.W., Le Grice, S.F. & Brown, P.O. Two-dimensional crystallization of histidine-tagged, HIV-1 reverse transcriptase promoted by a novel nickel-chelating lipid. *J Struct Biol* **113**, 117-23 (1994).
15. Hoffmeyer, S. *et al.* On unequal allelic expression of the neurofibromin gene in neurofibromatosis type 1. *Hum Mol Genet* **4**, 1267-1272 (1995).

6. Bibliography of publications and meeting abstracts

Publications

1. Li, Y., O'Connell, P., Breidenbach, H.H., Cawthon, R., Stevens, J, Xu, G., Neil, S., Robertson, M., White, R., and Viskochil, D. (1995). Genomic organization of the neurofibromatosis 1 gene (NF1). *Genomics* **25**, 9-18.
2. Purandare, S.M., Breidenbach, H.H., Li, Y., Zhu, X.L., Sawada, S., Neil, S. M., Brothman, A., White, R., Cawthon, R., and Viskochil, D. (1995). Identification of Neurofibromatosis 1 (*NF1*) homologous loci by direct sequencing, fluorescence *in situ* hybridization, and PCR amplification of somatic cell hybrids. *Genomics* **30**, 476-485.
3. Martinez, J.M., Breidenbach, H.H. & Cawthon, R.M. (1996). Long RT-PCR of the entire 8.5 Kb *NF1* Open Reading Frame and mutation detection on agarose gels. *Genome Research* **6**, 58-66.

Meeting abstracts

1. Breidenbach, H.H. & Cawthon, R. Identification Of *NF1* Mutations By A Protein Truncation Assay. 1994 FASEB Summer Research Conferences, Santa Cruz, California. Neurofibromatosis: July 9-14.
2. Cawthon, R. & Breidenbach, H.H. Identification Of *NF1* Mutations By A Protein Truncation Assay. 1994 Annual Meeting of the American Society for Human Genetics, Montreal, Quebec, Canada. October 18-22. Abstract published in *Am J Hum Genet* **55**, A216.

7. List of all personnel receiving pay from this grant

9/93 - 10/96: Richard Cawthon, P.I.
9/93 - 10/94: Heidi Huntsman Breidenbach, Technician.
10/94 - 10/95: Jennifer M. Martinez, Technician.
10/95 - 10/96: Xiufeng Wu, Technician.

GENOMICS 25, 9–18 (1995)

Genomic Organization of the Neurofibromatosis 1 Gene (*NF1*)

YING LI,* PETER O'CONNELL,*,[1] HEIDI HUNTSMAN BREIDENBACH,* RICHARD CAWTHON,* JEFF STEVENS,* GANGFENG XU,*,[2] SHANNON NEIL,† MARGARET ROBERTSON,* RAY WHITE,* AND DAVID VISKOCHIL†,[3]

*Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112; and †Department of Pediatrics, University of Utah, Salt Lake City, Utah 84112

Received June 22, 1994; revised September 23, 1994

Neurofibromatosis 1 maps to chromosome band 17q11.2, and the *NF1* locus has been partially characterized. Even though the full-length *NF1* cDNA has been sequenced, the complete genomic structure of the *NF1* gene has not been elucidated. The 5′ end of *NF1* is embedded in a CpG island containing a *Not*I restriction site, and the remainder of the gene lies in the adjacent 350-kb *Not*I fragment. In our efforts to develop a comprehensive screen for *NF1* mutations, we have isolated genomic DNA clones that together harbor the entire *NF1* cDNA sequence. We have identified all intron–exon boundaries of the coding region and established that it is composed of 59 exons. Furthermore, we have defined the 3′-untranslated region (3′-UTR) of the *NF1* gene; it spans approximately 3.5 kb of genomic DNA sequence and is continuous with the stop codon. Oligonucleotide primer pairs synthesized from exon-flanking DNA sequences were used in the polymerase chain reaction with cloned, chromosome 17-specific genomic DNA as template to amplify *NF1* exons 1 through 27b and the exon containing the 3′-UTR separately. This information should be useful for implementing a comprehensive *NF1* mutation screen using genomic DNA as template.

INTRODUCTION

Neurofibromatosis type 1 (NF1) afflicts approximately 1 in 3500 individuals worldwide. This autosomal dominant condition typically manifests café-au-lait macules in the first decade of life and cutaneous neurofibromas during adolescence. Approximately $\frac{1}{3}$ of afflicted individuals display assorted medical complications involving many organ systems, and they carry a small but increased risk for the development of malignancy. In addition to the highly pleiotropic nature of NF1, the clinical manifestations are quite variable, even among family members who carry the same disease allele. The *NF1* gene was cloned by the mapping approach (Cawthon *et al.*, 1990a; Viskochil *et al.*, 1990; Wallace *et al.*, 1990), and its cDNA has been sequenced (Cawthon *et al.*, 1990a; Wallace *et al.*, 1990; Xu *et al.*, 1990a; Marchuk *et al.*, 1991). The coding segment of *NF1* spans approximately 350 kb of genomic DNA, and the cDNA is 8454 bp long (Marchuk *et al.*, 1991), exclusive of two alternative splice forms that contain insertion exons of 54 (Cawthon *et al.*, 1990a) and 63 (Nishi *et al.*, 1991; Suzuki *et al.*, 1991) bp. *NF1* encodes neurofibromin, a 2818-amino-acid polypeptide with domains of homology to yeast IRA1 and IRA2 and to mammalian p120GAP (p21ras GTPase-activating Protein) (Buchberg *et al.*, 1990; Xu *et al.*, 1990a). The domain demonstrating homology to the catalytic domain of p120GAP activates the intrinsic GTPase of p21ras *in vitro* (Ballester *et al.*, 1990; Martin *et al.*, 1990; Xu *et al.*, 1990b). Other functions of neurofibromin have not been conclusively demonstrated.

A detailed understanding of gene structure at the *NF1* locus is paramount in evaluating the molecular basis of the high mutation rate of NF1, estimated to be one of the highest among human genetic disorders (approximately 1/10,000 gametes per generation) (Carey *et al.*, 1986). Even though the *NF1* cDNA sequence is known, precise identification of *NF1* mutations in patients has not been very fruitful; mutations have been defined in fewer than 20% of screened *NF1* patients (NNFF International NF1 Genetic Analysis Consortium, Fax: (617) 277-5933) and include translocations, deletions, insertions, and point mutations that are generally "inactivating." The identification of exon boundaries and determination of intronic sequences flanking these boundaries would enable investigators to adopt additional screening approaches to facilitate the further evaluation of mutations in the *NF1* gene.

The availability of cloned genomic DNA from the cosmid contig containing a portion of the *NF1* cDNA (Viskochil *et al.*, 1990) enabled Cawthon *et al.* (1990a)

Variable lengths of intronic sequence on both sides of *NF1* exons 2 through 27b have been deposited with GenBank under Accession Nos. U17656–U17690. Continuous DNA sequence spanning exons 28 through 49 have been deposited under Accession No. L05367.

[1] Present address: University of Texas Health Sciences Center-San Antonio, Department of Pathology, San Antonio, TX 78284.

[2] Present address: Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02114.

[3] To whom correspondence should be addressed at the Division of Medical Genetics, 413 MREB, University of Utah, Salt Lake City, UT 84112. Telephone: (801) 581-8943. Fax: (801) 585-5241.

0888-7543/95 $6.00

initially to identify nine exon boundaries with flanking intronic sequence. Subsequently, we used the same clones to define the boundaries of downstream exons (this report), and these have been confirmed by DNA sequence generated by Weiss *et al.* (1992). Finally, *NF1* cDNAs mapping further 5′ and extending centromeric of the cosmid contig were used to isolate genomic clones to generate exon boundaries in the 5′ portion of the gene. Based on preliminary results, an exon numbering system was adopted by consensus at the NNFF International Consortium on Gene Cloning and Gene Function for NF1 and NF2 (Oxford, UK, 1992), which redesignated exon 1 reported by Cawthon *et al.* (1990a) as exon 28; the exon containing the stop codon was designated exon 49. We report here the genomic organization of all *NF1* exons, and we present exon boundaries with flanking intronic sequences.

To complete the structural analysis of the *NF1* gene, we characterized the 3′ end of *NF1* cDNAs. None of the cDNA clones used originally to identify the *NF1* gene included polyadenylation tracts that typically mark the 3′ end of mRNA transcripts. Furthermore, the 11- to 13-kb size of human *NF1* mRNA on Northern analysis (Buchberg *et al.*, 1990; Viskochil *et al.*, 1990; Wallace *et al.*, 1990; Suzuki *et al.*, 1992) predicts that there is an additional 2.5 to 4 kb of noncoding sequence in the mature *NF1* transcript. One human cDNA clone from an oligo(dT)-primed library made from a basophilic leukemia cell line was reported to contain the 3′ end of *NF1* (Suzuki *et al.*, 1992); however, only a segment of that clone, pHN-2, was sequenced, and the 3′-untranslated region was not reported. Bernards *et al.* (1993) have cloned and sequenced mouse cDNAs that contain up to 3211 bp of DNA beyond the mouse *nf1* stop codon and have shown that the mouse 3′-UTR cDNA sequence is highly homologous to the human genomic sequence lying immediately downstream of its *NF1* stop codon. In this report we define the 3′ end of human *NF1*, and we demonstrate that the 3′-untranslated region of *NF1* mRNA from four different tissues is continuous with the stop codon in exon 49.

MATERIALS AND METHODS

Isolation of genomic clones. *NF1*-specific genomic clones were selected from a cosmid library that was constructed with DNA prepared from human peripheral blood leukocytes and inserted into the *Bam*HI restriction site of the pWE15 vector (Stratagene, La Jolla, CA). A premade P1-bacteriophage human genomic library (The DuPont–Merck Pharmaceutical Co., Wilmington, DE, DMPC-HFF#1 Series B) cloned into the *Bam*HI site in the pAd10sacBII vector (Pierce and Sternberg, 1992) provided a second source of genomic clones. High-density P1-bacteriophage libraries were screened as previously described (O'Connell *et al.*, 1990) using radioactive probes representing different *NF1* cDNA sequences. Hybridization on nylon membranes was carried out in 50% formamide, 5× SSC, 50 m*M* sodium phosphate (pH 6.5), 2× Denhardt's solution, 1% dextran sulfate, 0.1% SDS, and 100 μg of denatured total human DNA per milliliter for 14–16 h at 42°C. Filters were sequentially washed under increasingly stringent conditions, with the final wash in 0.1× SSC, 0.1% SDS at 55°C for 5 min for P1 bacteriophage libraries and at room temperature for 5 min for cosmid libraries. Positive clones were identified by autoradiography and purified by subsequent screenings using similar methodology. Isolated colonies were prepared by $CsCl_2$/ethidium bromide buoyant density centrifugation using standard protocols (Pierce and Sternberg, 1992; Sambrook *et al.*, 1989). Genomic clones were sized by agarose gel electrophoresis of *Eco*RI- and *Bam*HI-digested fragments both under standard conditions (Sambrook *et al.*, 1989) and using field-inversion methodology (Carle *et al.*, 1986). The overlapping ends of selected cosmid clones were mapped by Southern analysis using T7 and T3 polymerase-generated RNA probes synthesized from the ends of the clones.

cDNA screening. Premade Lambda ZAP cDNA libraries (Stratagene) were screened with PCR-labeled probes as previously described (Xu *et al.*, 1990a). Libraries constructed from four different tissue sources were screened: fetal brain (random and oligo(dT) primed, Stratagene, 936206), fetal retina from pooled tissue (oligo(dT) primed, Stratagene, 937202), muscle from an adult female (oligo(dT) primed, Stratagene, 936215), and adult heart (random and oligo(dT) primed, Stratagene, 936207). Isolated clones were excised according to the manufacturer's recommendations (Stratagene Protocol 200253), and the resultant *NF1* cDNA inserts in pBluescript SK(−) vectors were sized and sequenced by standard techniques.

Probe synthesis. PCR probes were synthesized by incorporating [α-^{32}P]dCTP into a PCR product (Saiki *et al.*, 1988). Template DNA was obtained from clones representing *NF1* cDNA between exons 2 and 27b. PCR conditions for probe synthesis were as follows: 1–10 ng template DNA, 1.2 m*M* dATP/dTTP/dGTP, 2.5 μ*M* dCTP, 10–15 μl of [α-^{32}P]dCTP (Amersham, Arlington Heights, IL; sp act, >6000 Ci/mmol), 4 m*M* $MgCl_2$, 10 m*M* Tris (pH 8.3), 50 m*M* KCl, 0.5 m*M* primers, and 2.5 units *Taq* polymerase, in a total volume of 50 μl. Routine thermal cycling included denaturation at 93°C, variable annealing temperatures, and extension at 72°C for 1 min for 20 cycles. The final step included heat denaturation at 93°C for 3 min, and the probe was placed on ice for at least 5 min before being added (40 μl) to the hybridization solution.

RNA probes from the ends of cosmid inserts were synthesized according to the manufacturer's instructions accompanying the T7 and T3 RNA polymerase transcription kit (Ambion, Austin, TX; No. 1326). Cosmid DNA (1–2 mg) digested with *Rsa*I served as template for probe synthesis using [α-^{32}P]UTP (Amersham, sp act, 800 Ci/mmol). The probes were routinely used directly in Southern analysis at template concentrations of 10 ng/ml hybridization solution.

Identification of intron–exon boundaries. Initial characterization of exon boundaries and flanking intronic sequences was performed in an empiric fashion by a PCR approach. Facing *NF1* cDNA oligonucleotides, tagged with universal primer sequence at the 5′ ends (universal (21m13)—GTAAAACGACGGCCAGT; reverse (m13RP1)—AACAGCTATGACCATG), were synthesized approximately 150 bp apart according to the cDNA sequence. These primers were used to amplify DNA isolated from genomic clones. PCR conditions varied for each primer pair; however, the routine reaction was based on the manufacturer's recommendations for the respective *Taq* polymerase, and the annealing temperature was set at approximately 5°C below the estimated T_m for 35 cycles. PCR products larger than the size predicted on the basis of cDNA sequence presumably carried an intron, and each PCR product was sequenced from both directions to identify divergence points from cDNA sequence. The presence of consensus splice junctions at sites where the sequence of a PCR product diverged from known *NF1* cDNA sequence identified the intron–exon boundaries. Initially, for those "introns" that did not amplify, boundaries were obtained by direct DNA sequencing of subclones. In this case, DNA from genomic clones was digested with *Sau*3A, subcloned in the *Bam*HI site of pBluescript, and selected with *NF1* cDNA probes. Plasmid DNA from positive clones was directly sequenced using either the universal or the reverse primer and T7 DNA polymerase (Sequenase Version 2.0, U.S. Biochemicals (USB), Cleveland, OH). Later in the project other techniques replaced the subcloning protocols. Those regions that could not be amplified by primer pairs 150 bp apart were subjected to either anchored PCR (Roux and Dhanarajan, 1990) or inverse PCR (Ochman *et al.*, 1988) techniques to obtain exon boundaries and flanking intronic se-

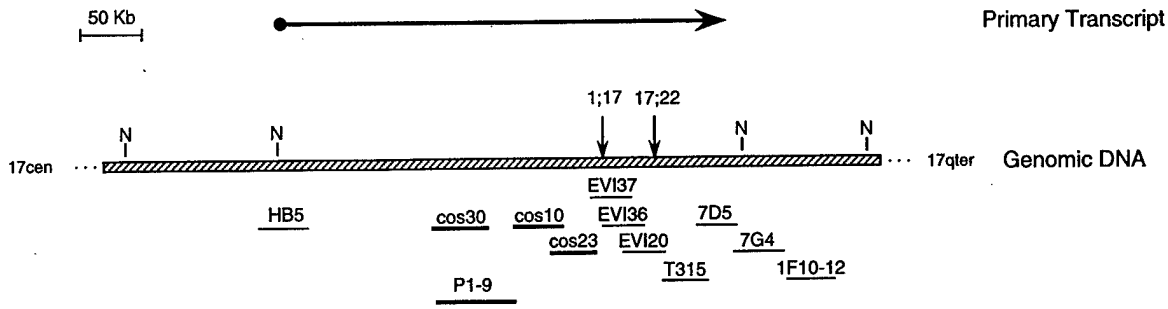


FIG. 1. The *NF1* locus. Genomic DNA is represented as a hatched bar; the length and transcriptional direction of *NF1* is depicted by the arrow above the region of genomic DNA that it spans. N denotes *Not*I restriction endonuclease cleavage sites (Fountain *et al.*, 1989; O'Connell *et al.*, 1989). Cosmid and P1 clones that cover all *NF1* exons are shown below the map. Clones not previously reported are indicated by thick bars. Breakpoints of two previously characterized chromosomal rearrangements, t(1;17) and t(17;22) (O'Connell *et al.*, 1990), are indicated by downward arrows above the genomic fragment.

quences. The PCR products were sequenced according to established methods of automated dideoxy sequencing using fluorescently labeled primers and either T7 or *Taq* polymerase.

DNA sequencing. Purified genomic DNAs from cosmid and P1 clones served as templates in routine PCR using *NF1*-specific primers tagged with 21m13 (upstream primer) and m13RP1 (downstream primer). The resulting PCR products were purified by Centricon-100 ultrafiltration (Amicon, Beverly, MA) and sequenced using a test-site protocol provided by Applied Biosystems, Inc. (Foster City, CA). This procedure involved performing dideoxy sequencing reactions with *Taq* polymerase in a thermal cycler, using fluorescently tagged M13 universal (UP) or reverse (RP) sequencing primers, for analysis on an Applied Biosystems Model 373A automated sequencer.

Subcloned genomic DNA in pBluescript was sequenced by the dideoxy method using either [α-^{32}P]dATP or [^{35}S]dATP (Amersham, Arlington Heights, IL) and a Sequenase II kit with UP and RP sequencing primers (USB). cDNA inserts in pBluescript vectors were sequenced with T7 polymerase (Sequenase Version 2.0, USB) using fluorescently tagged UP and RP sequencing primers in an ABI protocol (Cawthon *et al.*, 1990a). Internal sequence was obtained manually using *NF1*-specific primers with a sequencing kit (Sequenase Version 2.0, USB) and [α-^{32}P]dATP.

RESULTS

Physical map of the 5′ end of the NF1 locus. We had previously generated a cosmid contig now known to cover the portion of the *NF1* locus harboring exons 28 through 49 (O'Connell *et al.*, 1990; Viskochil *et al.*, 1990). The contig bounded by cosmids EVI36 and 7D5 is approximately 101 kb (Weiss *et al.*, 1992). The telomeric end of 7D5 maps 2.5 kb upstream of the telomeric *Not*I site of the approximate-size 350-kb *Not*I fragment (R. Weiss, University of Utah, pers. comm., 1994). To characterize the 5′ end of the *NF1* gene further, we used *NF1* cDNAs as probes under stringent hybridization conditions to isolate genomic clones mapping upstream of exon 28. As shown in Fig. 1, the newly defined genomic contig consists of cosmids 30, 10, and 23 and the P1-bacteriophage clone P1-9. An overlap of approximately 7.5 kb between the two contigs was established by Southern blot analysis using RNA probes from the T7 end of cosmid EVI37 and the T3 end of cos23 (data not shown). On the basis of Southern blot analysis, we estimate that the contig consisting of cos30, P1-9, cos10, and cos23 is approximately 100 kb long. Cosmid HB5 (Jorde *et al.*, 1993) maps to the centromeric end of the 350-kb *Not*I fragment, and it extends approximately 20 kb telomeric of the *Not*I site (Fig. 1). The difference between the estimated size of the *Not*I fragment and the amount of genomic DNA contained in clones mapping between the *Not*I sites is about 115 kb; this gap lies between clones HB5 and cos30.

Even though the cloned contig does not cover the entire gene, all of the *NF1* exons are represented in genomic clones. Exon 1 maps just 5′ to the centromeric *Not*I site (Marchuk *et al.*, 1992) and is contained in cosmid HB5 (data not shown). Exon 2 maps to cos30 and P1-9 by Southern analysis and by sequence identity of PCR amplification products. Thus, uncloned genomic DNA lying in the gap between HB5 and cos30 is contained within intron 1.

NF1 exon boundaries. The cloning of genomic DNA encompassing all known *NF1* exons provided substrate for the delineation of exon boundaries and flanking intronic sequences. Using genomic clones from the contig, four approaches enabled us to generate exon-selected template for sequencing: (1) direct PCR using facing exon-based oligonucleotides as primers; (2) subcloning of *Sau*3A-digested DNA; (3) inverse PCR with tail-to-tail, exon-based primer pairs (Ochman *et al.*, 1988); and (4) "anchored PCR" using *NF1* cDNA-specific primers with a universal primer to amplify DNA fragments that have been ligated to a restriction site-specific, universal anchor adapter primer (Roux and Dhanarajan, 1990). The exon boundaries depicted in Fig. 2 designate sites where *NF1* cDNA sequence diverged from genomic DNA sequence. Around these sites of divergence, the sequences were typical of splice junctions.

In the course of this work several introns were identified after the adoption of the *NF1* exon numbering system (NNFF International Consortium on Gene Cloning and Gene Function for NF1 and NF2). We have found that exon 4 is actually divided into 4a, 4b, and 4c; exon 10 is divided into 10a, 10b, and 10c; and exon 12 is divided into 12a and 12b. Exon 23 is also interrupted by an intron, but because the previous designation of the alternatively spliced insertion exon 23a was al-

FIG. 2. Locations of exon boundaries in relation to the *NF1*-encoded amino acid sequence (initiating methionine is 1). Downward arrowheads indicate exon boundaries, and the exon number is marked to the right of each arrowhead at the 5′ end of its respective exon. The exon boundaries between 27a and 27b and between 19a and 19b have also been determined by others (Martin-Gallardo *et al.,* 1992; P. Robinson, pers. comm., 1994).

ready firmly established, we elected to designate the two exons derived from exon 23 as 23-1 and 23-2. We have confirmed an unpublished report that exon 19 is divided into exon 19a and 19b (P. Robinson *et al.,* Institute of Medical Genetics, Humboldt University, Berlin, pers. comm., 1994). Finally, exon 27 is divided into 27a and 27b (Martin-Gallardo *et al.,* 1992).

NF1 exon-flanking sequences were derived from the following genomic clones (see Fig. 1): exon 1—cosmid HB5; exons 2 through 6—cosmid c30; exons 7 through 10b—bacteriophage P1-9; exons 10c through 23-1—cosmid c10; exons 23-2 through 27b—cosmid c23; and exons 28 through 49—cosmids cEVI36, cEVI20, cT315, and c7D5. Oligonucleotide primers synthesized from intronic sequences were used to PCR-amplify *NF1* exons using DNA from *NF1* genomic clones as template. The sequences of these primer pairs are presented in Table 1 along with the size of each PCR product generated (shown in parentheses under the number of the exon that it contains).

Sizes of the majority of *NF1* introns are shown in Table 2. The exact lengths of all introns from exons 28 through 49 are known from sequence analysis. The lengths of introns between exons upstream of exon 27b have been estimated by agarose gel electrophoresis of PCR amplification products using facing, exon-based primer pairs from adjacent exons with cloned *NF1* genomic DNA as template. Inability to amplify by PCR indicated that the intron was likely to be longer than 4 kb in length. Numerous introns have been fully sequenced and are marked with an asterisk in column 4 of Table 2.

Intron 10c harbors sequence from a cDNA, pHN-3, that was isolated from a human placenta library (Suzuki *et al.,* 1992). We found that the divergence point of pHN-3 with respect to *NF1* cDNA lies at an exon boundary and that the sequence of pHN-3 cDNA downstream of this boundary is continuous genomic sequence that extends 241 residues into intron 10c. Thus, failure to splice exon 10c to exon 11 could lead to alternative processing of the *NF1* message, resulting in a

TABLE 1

Intron Primers Used to Amplify *NF1* Exons

Exon (bp)	Sequence (5′ → 3′)	Exon (bp)	Sequence (5′ → 3′)
1	CAGACCCTCTCCTTGCCTCTT	18	AGAAGTTGTGTACGTTCTTTTCT
(438)	GGATGGAGGGTCGGAGGCTG	(367)	CTCCTTTCTACCAATAACCGC
2	TTTTAAGGATAAACTGTTTACGTG	19a	TCATGTCACTTAGGTTATCTGG
(221)	TCCCCAAAACACAGTAACCC	(267)	TGTAATTAAGTAGTTATAACTCTC
3	TTTCACTTTTCAGATGTGTGTTG	19b	CCTAAAGTTTATATCTGTTAATAAG[a]
(195)	AATTCCAAGCCTCCTACTTAG	(276)	TGGTGGGGGGCTTTATTTGC
4a	TTTGAAAATTTTCATAATAGAAAATGT	20[b]	CCACCCTGGCTGATTATCG
(415)	GAGGTCAAAGCTGCTGTGAG	(402)	TAATTTTTGCTTCTCTTACATGC
4b	CTCCTGGCCTCAAGTGGTC	21	AAATGAAAGTTTCATATAGAAATAC
(339)	TTATAAAATCCAGATTGGTGTTC	(374)	ATTTGCTATGTGCCAGGGAC
4c	TTTCCTAGCAGACAACTATCGA	22[b]	TGCTACTCTTTAGCTTCCTAC
(283)	CATCAAAAAAAAAAATTTTAATACCAG	(331)	CCTTAAAAGAAGACAATCAGCC
5	TGACTTGAGTGATAGTTTCACAT	23-1	TTTGTATCATTCATTTTGTGTGTA
(240)	AAAAAAAAATCAATCGTATCCTTA	(281)	AAAAACACGGTTCTATGTGAAAAG
6	CATGTTTATCTTTTAAAAATGTTGCC	23-2[b]	CTTAATGTCTGTATAAGAGTCTC
(301)	ATAATGGAAATAATTTTGCCCTCC	(268)	ACTTTAGATTAATAATGGTAATCTC
7	TGCTATAATATTAGCTACATCTGG	23a	AGCCAGAAATAGTATACATGATTGGGT
(373)	CCTATGAACTTATCAACGAAGAG	(447)	CTATTTTCTGCCAGAATTAGTAGA
8	TGTGCTGCTTCTGGCAACTG[a]	24[b]	TTGAACTCTTTGTTTTCATGTCTT
(318)	CTAGTCTTTCTGTTTATAAAGGAT[a]	(266)	GGAATTTAAGATAGCTAGATTATC
9	TTTGACCTCATTTGTATTACTGAG	25[b]	AATATAATAATTATATTTGGGAAGGT
(248)	AGAACCTTTTGAAACCAAGAGTG	(338)	GAAAATATTTGATTCAAACAGAGC
10a	ACGTAATTTTGTACTTTTTCTTCC	26	GCTTTGTCTAATGTCAAGTCAC
(232)	CAATAGAAAGGAGGTGAGATTC	(342)	TTAAACGGAGAGTGTTCACTATC
10b	CTTTAAAGTGATAGCTATTACC	27a	GTTACAAGTTAAAGAAATGTGTAG
(267)	TCTTGGCGATTCAGCTAAACC	(298)	CTAACAAGTGGCCTGGTGGCAAAC[c]
10c	CTTGGTACCCTTTAGCAGTCAC	27b	TTTATTGTTTATCCAATTATAGACTT
(379)	CCTTCTTTCTCCATGGAG	(296)	TCCTGTTAAGTCAACTGGGAAAAAC[c]
11	GTACTCCAGTGTTATGTTTACC	48a	ATCTAGTATCTAATTGTATTTCACC
(189)	TAAAGTTGAAATTTAAAAATTAAAGTAC	(246)	GCAGACTGAGCTTACAGGGAC
12a	AAACCTTACAAGAAAAACTAAGCT	49.1	CTGGGAGAAACAGGCTATAC
(303)	ATTACCATTCCAAATATTCTTCCA	(327)	TTTGAAACTGCCAGCGCT
12b	TTTCTAGTGAATCTCCTTCAAGT	49.2	GCAATGAAATTCAGTCCTGGAAGG
(382)	ATGAAATTTACCAAATTTCATTCAG	(449)	CCCACTTTCTTTGCAGTTGTTCTG
13	CACAGTTTATTGCATTGTTAGAT	49.3	AAGAAAGTGGGAGGTCAGGAA
(382)	GCCATGTGCTTTGAGGCAGAA	(678)	ACTTTGTTCACTGATTACCCTGC
14[b]	TCCTTTTGGGTGGAGCTTATC	49.4	AAACTCAGTCATCCAAACTAG
(285)	TATACTTGTAATATGCACGTATC	(776)	GAAACTCCCCTCTCCCCTAG
15	TGTGATCAGGAATAGCTTTTGAA	49.5	GTAGTGCTGTTTCATTAGAGG
(275)	TTAACAGATAAAAGTCAACTTTAC	(888)	ACATTTTAGTTACTTGGAAAGCTAG
16[b]	TGGATAAAGCATAATTTGTCAAGT	49.6	ACCTTGTTTAAAGAACTTCTGAC
(549)	TAGAGAAAGGTGAAAAATAAGAG	(717)	CACAAACTATTTACAGGTGAATATAC
17	CTCTGTGTGTTTAGATCAGTCA	49.7	CAGGTTTGTAGGTTTTAGG
(318)	TTTATCAATTACTACCAGTATCAG	(707)	GGGAAGAAAACAGAGCCGATAC

Note. Oligonucleotides synthesized from intronic sequences were used in routine PCR to amplify the expected size products as noted in parentheses below each exon number. Exon 49 including the 3′-UTR has been amplified in seven overlapping segments.

[a] Sequence derived from clones that map to *NF1*-homologous loci.

[b] Primer pairs that yield *NF1*-specific sequence using uncloned genomic DNA as template.

[c] Primer sequence from Martin-Gallardo *et al.* (1992).

shorter mRNA species that encodes a smaller isoform of neurofibromin.

NF1 3′-untranslated region. The original reports of *NF1* cDNAs failed to identify the 3′ end of the gene. The majority of published *NF1* cDNA sequences were derived from a human fetal-brain library that was oligo(dT) and random-primed for reverse transcription into cDNA (Cawthon *et al.,* 1990a; Xu *et al.,* 1990a; Marchuk *et al.,* 1991). Other 3′-end cDNAs were isolated from various libraries, including cauda equina and B lymphoblasts (Wallace *et al.,* 1990), pre-B lymphocytes (Bernards *et al.,* 1992), basophilic leukemia cells (Suzuki *et al.,* 1992), and fetal muscle, adult brain, and endothelial cells (Marchuk *et al.,* 1991); the longest extended 291 bases downstream of the stop codon (Bernards *et al.,* 1992). None of these *NF1* cDNAs contained a poly(A) tail, multiple clones were chimeric, and no clones terminated at the same site. Subsequently, we isolated two cDNA clones that matched genomic sequence downstream of the stop codon, FB12 and HB4

TABLE 2

Coding Exons of *NF1*

Exon No.	cDNA position	Size (bp)	Introns (kb)	Exon No.	cDNA position	Size (bp)	Introns (kb)
1	1	60	? (20–140)	(23a	4111	63	6.0)
2	61	144	3.10	24	4111	159	0.53*
3	205	84	?	25	4270	98	1.25
4a	289	195	?	26	4368	147	1.27
4b	484	103	2.0	27a	4515	147	?
4c	587	68	0.22*	27b	4662	111	45–50
5	655	76	0.80	28	4773	433	1.3
6	731	158	?	29	5206	341	2.7
7	889	174	0.40	30	5547	203	4.3
8	1063	123	?	31	5750	194	1.55
9	1186	75	4.0	32	5944	141	0.15
10a	1261	142	?	33	6085	280	0.40
10b	1393	135	4.0	34	6365	215	0.24
10c	1528	114	2.5	35	6580	62	0.15
11	1642	80	0.54	36	6642	115	0.57
12a	1722	124	?	37	6757	102	1.7
12b	1846	156	1.2	38	6859	141	2.4
13	2002	250	0.49	39	7000	127	6.0
14	2252	74	0.23*	40	7127	132	0.93
15	2326	84	1.3	41	7259	136	2.0
16	2410	441	0.38*	42	7395	158	4.0
17	2851	140	0.28*	43	7553	123	0.35
18	2991	123	0.46*	44	7676	131	0.18
19a	3114	84	1.2	45	7807	101	1.1
19b	3198	117	0.55	46	7908	143	0.35
20	3315	182	0.12*	47	8051	47	1.4
21	3497	212	2.2	48	8098	217	6.5
22	3709	162	0.14*	(48a	8315	54	6.7)
23-1	3871	104	?	49	8315	153	
23-2	3975	136	4.0		— 3′ Noncoding —		

Note. Intron lengths were determined either by PCR amplification using facing exon primers from adjacent exons or by complete sequence between exons (denoted by an asterisk). Question marks denote inability to amplify by PCR across the intron; such introns are likely to be larger than 4 kb.

(Fig. 3). FB12 was selected with a PCR probe from the 5′ end of exon 49, and H4B was selected with a PCR probe from exon 48a. FB12 ends at base 8749 (cDNA numbering system from Marchuk *et al.,* 1991), which is the same termination site as the cDNA clone designated Nalm6 (Bernards *et al.,* 1992), and H4B terminates at base 8926, respectively, 291 and 474 bases downstream of the stop codon. Based on Northern analysis of mRNA species that are approximately 11 and 13 kb in length (Viskochil *et al.,* 1990; Wallace *et al.,* 1990; Suzuki *et al.,* 1992), one would predict a 3′-UTR of at least 2.5 kb; therefore, it seemed unlikely that H4B and FB12 represented the 3′ end of the gene. The observation that the mouse *nf1* cDNA is highly homologous to human *NF1* genomic DNA downstream of the stop codon (Bernards *et al.,* 1993) suggested that genomic probes downstream of the stop codon could identify cDNAs that harbor 3′-UTR sequences. A PCR probe representing bases 86168 to 87104 in the genomic sequence (Weiss *et al.,* 1992) identified two independent clones from a fetal-retina cDNA library, FR5 and FR11, and one clone, M13, from an adult-muscle cDNA library; furthermore, a probe from bases 88724 through 89159 (Weiss *et al.,* 1992) identified a clone, FB21D, from a fetal-brain cDNA library. The sequences from the ends of the cDNA inserts matched the genomic sequence downstream of the *NF1* stop codon, and the insert sizes were those predicted based on the uninterrupted genomic sequence. The cDNAs that map to the 3′ end of *NF1* are schematically represented in Fig. 3. Clone FR5 is the longest, covering approximately 3.45 kb; it begins with base 8508 (Marchuck *et al.,* 1991), 53 bp downstream from the stop codon (base 8455), and

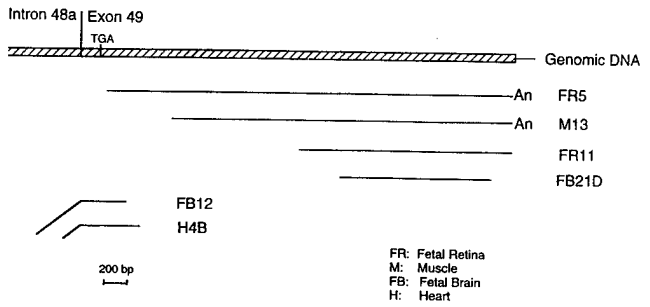


FIG. 3. The 3′-untranslated region of *NF1*. Genomic DNA is represented as a hatched bar, and the boundary between intron 48a and exon 49 is marked. The stop codon 143 bp downstream of the boundary is represented as a TGA above the genomic sequence. Six cDNAs mapping to the 3′-UTR genomic sequence are denoted by horizontal lines; clones FB12 and H4B contain upstream cDNA sequence denoted by an angle at the intron–exon boundary.

terminates in a poly(A) tract at base 11954 in the cDNA sequence, which corresponds to base 89095 in the genomic sequence (Weiss *et al.,* 1992). This clone identifies the *NF1* 3′-UTR as continuous with the stop codon and suggests that splicing does not occur in the 3′-untranslated region.

Both cDNA clones FR5 and M13 terminate at the same base with the addition of poly(A) tails of 18 residues added to an identical site in the genomic sequence. The genomic sequence spanning the polyadenylation site, TTAAGTAAAATGTAAATTCAATCT, contains a potential polyadenylation signal, AGUAAA, an uncommon sequence variant that is only $\frac{1}{3}$ as efficient as AAUAAA in polyadenylation activity (Wickens, 1990). There are two adenosine residues in the genomic sequence at the site of poly(A) addition; therefore, 18 residues could be attached to the boldfaced C in the same sequence shown (position 11954), or 16 residues could be attached to the boldfaced A (position 11956). Clone FR11 extends to base 11957 (additional T residue), but it does not contain a poly(A) tail. Neither does clone FB21D, which covers bases 10603 through 11838. Two sequence differences are notable between the 3′-UTR cDNAs and cosmid 7D5, which served as the template for the genomic sequence. Clones FR5, FR11, and M13 contain a 9-bp insertion (underlined) at base 10408 shown in the following sequence: CAGTGCCAAGGATGCCAAGCTGCCACC. Seven of the nine bases represent a duplication of an immediate upstream sequence. Clone FB21D carries a 7-bp deletion of a tandem duplication at base 11346, underlined in the following sequence derived from the cosmid clone 7D5: CAGTTTTGTTTGGGTTTTTTGTTTTTTT-GTTTTTGTTTCTA. The possibility that these sequence variants may represent useful polymorphisms is currently being investigated.

The contig of *NF1* cDNA clones shown in Fig. 3 indicates that the stop codon and the 3.5-kb 3′-untranslated region lie within the same *NF1* exon in genomic DNA. This observation establishes the 3′ end of the gene either at base 11954 or at base 11956 in the cDNA sequence and maps the end of *NF1* approximately 15 kb upstream of the telomeric *Not*I site.

DISCUSSION

The cDNAs described in this report identify the 3′ end of the *NF1* mRNA and establish the telomeric boundary of the *NF1* gene. *NF1* spans all but approximately 15 kb of an estimated 350-kb *Not*I fragment; exons 2 through 47, which includes the stop codon and the 3′-UTR, lie within this single fragment. Exon 1, which contains 5′-untranslated sequence in addition to coding sequence, maps to an adjacent 120-kb *Not*I fragment and lies immediately downstream of a functional promoter of transcription (G. Xu and R. White, unpublished data). The intron between exon 1 and exon 2 could be as large as 140 kb; 20 kb of genomic DNA downstream of exon 1 (represented in cosmid HB5) plus an estimated 115 kb of uncloned genomic DNA (this gap represents the portion of the estimated 350-kb *Not*I fragment not accounted for in the cloned contig) plus a maximum of 5 kb depending on the position of exon 2 within the end fragment of cosmid c30, the most 5′ genomic clone in the contig. Exons 2 through 27b are contained in approximately 100 kb of cloned genomic sequence, and exons 28 through 49 are contained in approximately 52 kb of sequenced genomic DNA. Like intron 1, intron 27b is also relatively large, approximately 45 to 50 kb, and it harbors three embedded genes, *EVI2A* (Cawthon *et al.,* 1990b), *EVI2B* (Cawthon *et al.,* 1991), and *OMGP* (Viskochil *et al.,* 1991). Intron 27b also contains two variable-repeat sites, separated by approximately 31 kb of genomic sequence, which have been adapted by clinical laboratories for linkage analysis (Lazaro *et al.,* 1993; K. Ward, University of Utah, pers. comm., 1994).

The role that the two large introns may play in *NF1* expression is unknown. The genes embedded in intron 27b are likely to place some constraints on the ubiquitous expression of *NF1*, although no transcriptional interference has been demonstrated. Even though an accurate ascertainment of the size of intron 1 will not be possible until the cloned contig is linked to cosmid HB5, its estimated size of 140 kb suggests that intron 1 could harbor regulatory elements that influence *NF1* expression in some way. In view of the presence of genes within intron 27b, the possibility that embedded genes also lie within this relatively large intron should be considered.

Knowledge of *NF1* structure enabled us to locate other landmarks in the gene. Amino acid homology with yeast Ira1 and Ira2 covers encoded exons 16 through 40. The homology with bovine P120GAP extends from exon 21 through 27a. Two known *NF1* translocation breakpoints t(1;17) and t(17;22) map to introns 27b and 31, respectively. The *AK3* pseudogene (Xu *et al.,* 1992) maps to intron 39. Finally, the two known alternative splice exons, 23a (Nishi *et al.,* 1991; Suzuki *et al.,* 1991) and 48a (Cawthon *et al.,* 1990a; Gutmann *et al.,* 1993), respectively, map to the 10-kb intron 23-2 and the 13.2-kb intron 48.

The two *NF1* cDNA clones that end in poly(A) tails define the 3′-UTR and establish the full length of the mRNA transcript as approximately 12 kb. There is evidence that mouse *nf1* cDNAs demonstrate alternative polyadenylation (Bernards *et al.,* 1993). Given that human *NF1* 3′-UTR contains other potential polyadenylation signals—AUUAAA at positions 9919, 10000, and 10086 and AUUAAA at positions 9140 and 9979—we evaluated all isolated cDNA clones for alternative polyadenylation. Although the cluster of polyadenylation signals between bases 9919 and 10086 could explain the approximate 2-kb difference between an 11- and a 13-kb mRNA species identified by Northern blot analysis (Suzuki *et al.,* 1992), we have not identified human *NF1* cDNAs that end within 20 bases downstream of

those motifs. Therefore, the basis of the size difference remains unknown.

Several sequence motifs identified in 3′-UTRs from other genes are known to affect mRNA stability and gene regulation (Jackson, 1993; Rastinejad and Blau, 1993); therefore, we speculate that the *NF1* 3′-UTR may play some role in the variability of clinical expression of NF1 by influencing *NF1* mRNA levels in a stochastic manner. However, there are few motifs present that could provide clues toward understanding its role in mRNA processing. The *NF1* 3′-UTR contains no UA-rich mRNA instability determinants (consensus UUAUUUAU), although several adenylation control elements are present (ACEs—consensus UUUUUUAU or UUUUAAU) (Jackson, 1993), of which the one nearest to the AAGUAA polyadenylation signal lies 460 bases upstream. A search of the nucleic acid sequence databases identified two homologies with 3′-UTRs from other genes; one is a short GT-rich segment in β-actin 3′-UTR, and another is a 100-bp stretch in the α-tropomyosin 3′-UTR. The functional significance of these homologies and motifs has not been tested.

Boundaries and splice junctions have been identified for all of the known *NF1* exons. Examination of the intronic splice sites reveals general adherence to the GT–AG rule (Shapiro and Senapathy, 1987). The 5′-splice sites are invariant for GT, although 7 introns lack the consensus purine residue at the third position. Likewise, the 3′ splice sites are invariant for AG, and only introns 25 and 36 have a purine instead of a pyrimidine in the preceding position. Notably, the 3′ splice junction between exons 10c and 11 lacks a polypyrimidine tract immediately upstream of the fourth position; instead, it has a stretch of 7 adenosine residues preceded by 13 pyrimidine residues. This sequence structure may confer a regulatory function for alternative processing leading to the 2.9-kb mRNA isoform from human placenta that terminates within exon 10c (Suzuki *et al.*, 1992).

Knowledge of the *NF1* exon boundaries and flanking sequence enabled us to design intron-based primer pairs to amplify individual exons from genomic DNA template. Using these specific primers in PCR amplifications, we have been able to determine the exon composition of genomic clones, to order the clones, and to map partially the regions of overlap within the contig. All of the known *NF1* exons are contained in either P1 or cosmid clones of genomic DNA. Sequence analysis of PCR products synthesized from the genomic clones shown in Fig. 1 revealed exon sequences identical to *NF1* cDNA sequence. Several other cosmid and P1 clones originally identified with *NF1* exon-based probes upstream of exon 28 were amplified by PCR using several primer pairs; however, the DNA sequences of exon regions from these PCR products are not identical to *NF1* cDNA sequence (data not shown). These clones map by fluorescence *in situ* hybridization to other chromosomes (D. Viskochil, unpublished data), and we interpret this to mean that they represent distant loci related to *NF1*. *NF1*-homologous loci have been described by others (Legius *et al.*, 1992; Gasparini *et al.*, 1993; Cummings *et al.*, 1993); the best characterized locus maps to chromosome 15. It shares about 90% homology with *NF1* within the intronic regions where sequences are known (Y. Li and R. Cawthon, unpublished data). Although a high degree of sequence homology exists between exons in the *NF1* GAP-related domain and the corresponding portion of the chromosome 15 locus, we have noted some frame-shifting deletions in the latter. Therefore, we conclude that even if it is transcribed, the chromosome 15 homologue does not encode a functional GAP-related protein. The identification of RT-PCR products where sequences are homologous, but not identical, to *NF1* exons (Cummings *et al.*, 1993) suggests that at least some *NF1*-related loci may be transcriptionally active. We have not fully evaluated the possibility that the chromosome 15 locus, or the other *NF1*-homologous loci, is expressed either in various tissues or in a developmentally regulated fashion.

We have synthesized oligonucleotide primers based on flanking intronic sequences for 56 of the 59 *NF1* exons and have demonstrated PCR amplification of a product of the expected size from genomic clones for each of them; primer pairs for exons 37, 38, and 39 remain to be synthesized. However, amplification by PCR of *NF1*-homologous loci when using oligonucleotide primers that are derived from intronic sequence from *bona fide NF1*-specific genomic clones complicates efforts to screen genomic DNA for *NF1* mutations by the direct-sequencing approach. This approach may require knowledge of intronic DNA sequences derived from cloned *NF1*-homologous loci before primer pairs can be designed to amplify *NF1* sequence specifically from genomic DNA. Using this information, we have synthesized seven primer pairs (superscript b in Table 1) that yield *NF1*-specific sequence using uncloned genomic DNA as template. The development of intron-based, exon-specific pairs of primers to amplify each *NF1* exon from genomic DNA is a logical first step toward implementation of a comprehensive, DNA-based screening protocol for *NF1* mutations.

ACKNOWLEDGMENTS

We acknowledge the technical expertise of Melanie Culver, primer synthesis by Edward Meenan, DNA sequencing by Paige Bradley, and Rick Lifton for providing the cosmid library from human peripheral blood leukocytes. We thank Ruth Foltz for her help in preparing the manuscript. R.C. was supported by grant awards from the National Neurofibromatosis Foundation (NNFF) and the National Cancer Institute (R55CA57511-01). P.O.C. was supported by Grant PO1 HG00470 from NIH. D.V. was supported as a clinical investigator by the National Institute of Neurological Disorders and Stroke (K08 NS01492), U.S. Department of Defense (DAMD 17-93-J-3070), Rocky Mountain Center for the Biology of Development under the Childrens Health Research Center, National Institute of Child Health, the NIH (5P30 HD27827), and the NNFF.

REFERENCES

Ballester, R., Marchuk, D., Boguski, M., Saulino, A., Letcher, R., Wigler, M., and Collins, F. (1990). The *NF1* locus encodes a protein functionally related to mammalian GAP and yeast *IRA* proteins. *Cell* **63:** 851–859.

Bernards, A., Haase, V., Murthy, A., Menon, A., Hannigan, G., and Gusella, J. (1992). Complete human *NF1* cDNA sequence: Two alternatively spliced mRNAs and absence of expression in a neuroblastoma cell line. *DNA Cell Biol.* **11:** 727–734.

Bernards, A., Snijders, A., Hannigan, G., Murthy, A., and Gusella, J. (1993). Mouse neurofibromatosis type 1 cDNA sequence reveals high degree of conservation of both coding and non-coding mRNA segments. *Hum. Mol. Genet.* **2:** 645–650.

Buchberg, A. M., Cleveland, L. S., Jenkins, N. A., and Copeland, N. G. (1990). Sequence homology shared by neurofibromatosis type-1 gene and *IRA-1* and *IRA-2* negative regulators of the *RAS* cyclic AMP pathway. *Nature* **347:** 291–294.

Carey, J., Baty, B., Johnson, J., Morrison, T., Skolnick, M., and Kivlin, J. (1986). The genetic aspects of neurofibromatosis. *Ann. N.Y. Acad. Sci.* **486:** 45–56.

Carle, G., Frank, M., and Olson, M. (1986). Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. *Science* **232:** 65–68.

Cawthon, R. M., Weiss, R., Xu, G., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P., and White, R. (1990a). A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell* **62:** 193–201.

Cawthon, R., O'Connell, P., Buchberg, A., Viskochil, D., Weiss, R., Culver, M., Stevens, J., Jenkins, N., Copeland, N., and White, R. (1990b). Identification and characterization of transcripts from the neurofibromatosis 1 region: The sequence and genomic structure of *EVI2* and mapping of other transcripts. *Genomics* **7:** 555–565.

Cawthon, R., Andersen, L., Buchberg, A., Xu, G., O'Connell, P., Viskochil, D., Weiss, R., Wallace, M., Marchuk, D., Culver, M., Stevens, J., Jenkins, N., Copeland, N., Collins, F., and White, R. (1991). cDNA sequence and genomic structure of *EVI2B,* a gene lying within an intron of the neurofibromatosis type 1 gene. *Genomics* **9:** 446–460.

Cummings, L., Glatfelter, A., and Marchuk, D. (1993). *NF1*-related loci on chromosomes 2, 12, 14, 15, 20, 21, and 22: A potential role for gene conversion in the high spontaneous mutation rate in NF1? *Am. J. Hum. Genet.* **53:** 672A.

Fountain, J., Wallace, M., Bruce, M., Seizinger, B., Menon, A., Gusella, A., Michels, V., Schmidt, M., Dewald, G., and Collins, F. (1989). Physical mapping of a translocation breakpoint in neurofibromatosis. *Science* **244:** 1085–1087.

Gasparini, P., Grifa, A., Origone, P., Coviello, D., Antonacci, R., and Rocchi, M. (1993). Detection of a neurofibromatosis type 1 (NF-1) sequence by PCR: Implications for the diagnosis and screening of genetic diseases. *Mol. Cell. Probes* **7:** 415–418.

Gutmann, D., Andersen, L., Cole, J., Swaroop, M., and Collins, F. (1993). An alternatively-spliced mRNA in the carboxy terminus of the neurofibromatosis type 1 (*NF1*) gene is expressed in muscle. *Hum. Mol. Genet.* **2:** 989–992.

Jackson, R. (1993). Cytoplasmic regulation of mRNA function: The importance of the 3′ untranslated region. *Cell* **74:** 9–14.

Jorde, L., Watkins, W., Viskochil, D., O'Connell, P., and Ward, K. (1993). Linkage disequilibrium in the neurofibromatosis 1 region: Implications for gene mapping. *Am. J. Hum. Genet.* **53:** 1038–1050.

Lazaro, C., Gaona, A., Ravella, A., Volpini, V., Casals, T., Fuentes, J.-J., and Estivill, X. (1993). Novel alleles, hemizygosity and deletions at an *Alu*-repeat within the neurofibromatosis type 1 gene. *Hum. Mol. Genet.* **2:** 725–730.

Legius, E., Marchuk, D., Hall, B., Anderson, L., Wallace, M., Collins, F., and Glover, T. (1992). NF-1 related locus on chromosome 15. *Genomics* **13:** 1316–1318.

Marchuk, D., Saulino, A., Tavakkol, R., Swaroop, M., Wallace, M., Andersen, L. Mitchell, A. Gutmann, D., Boguski, M., and Collins, F. (1991). cDNA cloning of the type 1 neurofibromatosis gene: Complete sequence of the *NF1* gene product. *Genomics* **11:** 931–940.

Marchuk, D., Tavakkol, R., Wallace, M., Brownstein, B., Taillon-Miller, P., Fong, C.-T., Legius, E., Andersen, L., Glover, T., and Collins, F. (1992). A yeast artificial chromosome contig encompassing the type 1 neurofibromatosis gene. *Genomics* **13:** 672–680.

Martin, G., Viskochil, D., Bollag, G., McCabe, P., Crosier, W., Haubruck, H., Conroy, L., Clark, R., O'Connell, P., Cawthon, R., Innis, M., and McCormick, F. (1990). The GAP-related domain of the NF1 gene product interacts with ras p21. *Cell* **63:** 843–849.

Martin-Gallardo, A., Marchuk, D., Gocayne, J., Kerlavage, A., McCombie, R., Venter, J., and Collins, F. (1992). Sequencing and analysis of genomic fragments from the *NF1* locus. *DNA Sequence—DNA Sequencing Mapping* **3:** 237–243.

Nishi, T., Lee, P., Oka, K., Levi, V., Tanase, S., Morino, Y., and Saya, H. (1991). Differential expression of two types of the neurofibromatosis type 1 (NF1) gene transcripts related to neuronal differentiation. *Oncogene* **6:** 1555–1559.

Ochman, H., Gerber, A., and Hartl, D. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* **120:** 621–623.

O'Connell, P., Leach, R., Cawthon, R., Culver, M., Stevens, J., Viskochil, D., Fournier, R., Rich, D., Ledbetter, D., and White, R. (1989). Two NF1 translocations map within a 600-kilobase segment of 17q11.2. *Science* **244:** 1087–1088.

O'Connell, P., Viskochil, D., Buchberg, A., Fountain, J., Cawthon, R., Culver, M., Stevens, J., Rich, D., Ledbetter, D., Wallace, M., Carey, J., Jenkins, N., Copeland, N., Collins, F., and White, R. (1990). The human homologue of murine *evi-2* lies between two translocation breakpoints associated with von Recklinghausen neurofibromatosis. *Genomics* **7:** 547–554.

Pierce, J., and Sternberg, N. (1992). Using bacteriophage P1 system to clone high molecular weight genomic DNA. *Methods Enzymol.* **216:** 549–574.

Rastinejad, F., and Blau, H. (1993). Genetic complementation reveals a novel regulatory role for 3′ untranslated regions in growth and differentiation. *Cell* **72:** 903–917.

Roux, K., and Dhanarajan, P. (1990). A strategy for single site PCR amplification of dsDNA: Priming digested cloned or genomic DNA from an anchor-modified restriction site and a short internal sequence. *BioTechniques* **8:** 48–57.

Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K., and Erlich, H. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shapiro, M., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **17:** 7155–7175.

Suzuki, Y., Suzuki, H., Kayama, T., Yoshimoto, T., and Shibahara, S. (1991). Brain tumors predominantly express the neurofibromatosis type 1 gene transcripts containing the 63 base insert in the region coding for GTPase activating protein-related domain. *Biochem. Biophys. Res. Commun.* **181:** 955–961.

Suzuki, H., Takahashi, K., Kubota, Y., and Shibahara, S. (1992). Molecular cloning of a cDNA coding for neurofibromatosis type 1 protein isoform lacking the domain related to ras GTPase-activating protein. *Biochem. Biophys. Res. Commun.* **187:** 984–990.

Viskochil, D., Buchberg, A., Xu, G., Cawthon, R., Stevens, J., Wolff, R., Culver, M., Carey, J., Copeland, N., Jenkins, N., and White, R. (1990). Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell* **62:** 187–192.

Viskochil, D., Cawthon, R., O'Connell, P., Xu, G., Stevens, J., Culver, M., Carey, J., and White, R. (1991). The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene. *Mol. Cell. Biol.* **11:** 906–912.

Wallace, M., Marchuk, D., Anderson, L., Letcher, R., Odeh, H., Saulino, A., Fountain, J., Brereton, A., Nicholson, J., Mitchell, A., Brownstein, B., and Collins, F. (1990). Type 1 neurofibromatosis gene: Identification of a large transcript disrupted in three patients. *Science* **249:** 182–186.

Weiss, R., Dunn, D., DiSera, L., Wheatley, W., Kimball, A., Rote, C., Cherry, J., Duval, B., Lee, R., Ferguson, M., and Gesteland, R. (1992). The human neurofibromatosis type 1 locus: Genomic sequence of the 3′ region, 1-100849. GenBank Accession No. L05367.

Wickens, M. (1990). How the messenger got its tail: Addition of poly(A) in the nucleus. *Trends Biochem. Sci.* **15:** 277–281.

Xu, G., O'Connell, P., Viskochil, D., Cawthon, R., Robertson, M., Culver, M., Dunn, D., Stevens, J., Gesteland, R., White, R., and Weiss, R. (1990a). The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell* **62:** 599–608.

Xu, G., Lin, B., Tanaka, K., Dunn, D., and Wood, D. (1990b). The catalytic domain of the NF1 gene product stimulates ras GTPase and complements IRA mutants of *S. cerevisiae*. *Cell* **63:** 835–841.

Xu, G., O'Connell, P., Stevens, J., and White, R. (1992). Characterization of human adenylate kinase 3 (*AK3*) cDNA and mapping of the *AK3* pseudogene to an intron of the *NF1* gene. *Genomics* **13:** 537–542.

GENOMICS 30, 476–485 (1995)

Identification of Neurofibromatosis 1 (*NF1*) Homologous Loci by Direct Sequencing, Fluorescence *in Situ* Hybridization, and PCR Amplification of Somatic Cell Hybrids

SMITA M. PURANDARE,* HEIDI HUNTSMAN BREIDENBACH,† YING LI,† XIAO LIN ZHU,† SHUN'ICHI SAWADA,‡ SHANNON M. NEIL,* ARTHUR BROTHMAN,* RAY WHITE,§ RICHARD CAWTHON,† AND DAVID VISKOCHIL*,[1]

*Departments of *Pediatrics and †Human Genetics and §Oncological Sciences and Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112; and ‡Department of Dermatology, The Jikei University School of Medicine, Tokyo 105, Japan*

Received April 3, 1995; accepted July 27, 1995

Using fluorescence *in situ* hybridization (FISH), we have identified seven *NF1*-related loci, two separate loci on chromosome 2, at bands 2q21 and 2q33–q34, and one locus each on five other chromosomes at bands 14q11.2, 15q11.2, 18p11.2, 21q11.2–q21, and 22q11.2. Application of PCR using *NF1* primer pairs and genomic DNA from somatic cell hybrids confirmed the above loci, identified additional loci on chromosomes 12 and 15, and showed that the various loci do not share homology beyond *NF1* exon 27b. Sequenced PCR products representing segments corresponding to *NF1* exons from these loci demonstrated greater than 95% sequence identity with the *NF1* locus. We used sequence differences between *bona fide NF1* and *NF1*-homologous loci to strategically design primer sets to specifically amplify 30 of 36 exons within the 5′ end of the *NF1* gene. These developments have facilitated mutation analysis at the *NF1* locus using genomic DNA as template.

INTRODUCTION

Neurofibromatosis 1 is an autosomal dominant disorder that afflicts approximately 1 in 3500 individuals worldwide (Crowe *et al.,* 1956; Riccardi, 1992). It manifests in an age-related fashion and is characterized mainly by cafe au lait spots, neurofibromas, and Lisch nodules of the iris. The *NF1* gene maps to chromosome 17q11.2 (Barker *et al.,* 1987), and, even though it was cloned in 1990 (Cawthon *et al.,* 1990; Viskochil *et al.,* 1990; Wallace *et al.,* 1990), its complete genomic organization was reported only recently (Li *et al.,* 1995). It is a large and complex gene that spans approximately 335 kb of genomic DNA and encodes a protein (neurofibromin) of 2818 amino acids (Marchuk *et al.,* 1991). The intron/exon boundaries of the coding region were recently characterized, and it was established that *NF1* is composed of 60 exons, inclusive of three alternatively spliced insertion exons (Li *et al.,* 1995; Danglot *et al.,* 1995). The *NF1* promoter region (Hajra *et al.,* 1994) and the 3′ untranslated region, which spans 3.5 kb of genomic DNA (Li *et al.,* 1995), have also only recently been characterized. *NF1* functions in part as a tumor suppressor gene, whereby inactivating mutations lead to loss of neurofibromin function associated with typical NF1-related tumors as well as other malignancies (Andersen *et al.,* 1993; Basu *et al.,* 1992; DeClue *et al.,* 1992; Johnson *et al.,* 1993; Legius *et al.,* 1993; Li *et al.,* 1992; The *et al.,* 1993; Shannon *et al.,* 1994; Xu *et al.,* 1992). This function is mediated, at least in part, by the GAP (GTPase activating protein)-related domain that, *in vitro,* stimulates hydrolysis of p21ras-GTP (Ballester *et al.,* 1990; Martin *et al.,* 1990; Xu *et al.,* 1990a). Thus, inactivating *NF1* mutations could play a significant role in tumorigenesis via altered signal transduction through p21ras.

Reliable identification of *NF1* mutations by PCR-based screening of *NF1* exons in genomic DNA would be useful for studies of NF1 disease and for analysis of the role of somatic *NF1* mutations both in sporadic tumorigenesis and in the variability of its clinical expression. The large size of the gene, high mutation rate, and the predominance of small deletions/insertions and point mutations have complicated the detection of *NF1* mutations. In addition, other loci have been observed to coamplify with the *NF1* gene in some PCR reactions when performed with primer pairs that lie upstream of exon 28, including segments that encode the catalytic domain of neurofibromin. An *NF1* related locus on chro-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. U35684–U35699.

[1] To whom correspondence should be addressed at the Division of Medical Genetics, 413 MREB, University of Utah, Salt Lake City, UT 84112. Telephone: (801) 581-8943. Fax: (801) 585-5241.

0888-7543/95 $12.00

mosome 15 (GenBank Accession No. M84131) was reported by Legius *et al.* (1992), and a different locus on chromosome 15 that mapped to band 15q24–qter (NF1HHS; EMBL Accession No. X72619) was detected by Gasparini *et al.* (1993). *NF1*-homologous loci on chromosomes 14 and 22 were reported by Marchuk *et al.* (1991), and an *NF1*-related locus on chromosome 21 at 21q11.2 was reported by Suzuki *et al.* (1994). Other *NF1*-related loci, reported in abstract form, have been identified on chromosomes 2, 12, and 20 (Cummings *et al.,* 1993). The extent of homologies between *NF1* and the various *NF1*-related loci has not been elucidated.

In the process of sequencing genomic DNA clones both to determine *NF1* exon boundaries and to develop intron-based PCR primers for individual *NF1* exons, we identified exonic sequences from some of our genomic clones that did not match *NF1* cDNA sequence. In addition, sequence of several PCR products generated from total human genomic DNA from non-NF1 controls revealed genomic sequence that was clearly derived from more than one locus. This strongly suggested the presence of additional loci in the genome. We used the genomic clones as probes in FISH (fluorescence *in situ* hybridization) analysis to identify seven *NF1*-related loci, and we used PCR of monochromosomal somatic cell hybrids to identify two additional loci. To improve *NF1* mutation analysis using total genomic DNA as a template, we have designed a number of oligonucleotide primer pairs to amplify specifically from the *NF1* locus on chromosome 17 by PCR. Implementation of these primer pairs would simplify germline and somatic mutation detection using PCR-based methods and genomic DNA as a template.

MATERIALS AND METHODS

Isolation of NF1 genomic clones. P-1 bacteriophage clones containing human genomic DNA were identified with *NF1* exon probes by a high-density screening protocol (Li *et al.,* 1995). Pooled probes containing sequence from exons 8 through 10c, 18 through 21, and 26 through 27b were labeled by using *NF1* cDNA as template and exon-based oligonucleotides as primers. Hybridization was carried out under standard high-stringency conditions as previously described (Li *et al.,* 1995). Multiple P-1 bacteriophage clones were identified, and they were partially mapped to the *NF1* locus by their ability to serve as template for PCR under standard conditions using facing primers encompassing specific *NF1* exons. Cosmid clones carrying human genomic DNA prepared from peripheral blood leukocytes in the pWE15 vector (Stratagene, San Diego, CA) were isolated by standard techniques as previously described (Li *et al.,* 1995).

DNA sequencing. *NF1*-specific primers tagged with 21m13 (upstream primer) and m13RPI (downstream primer) sequences were used for PCR amplification of the purified genomic DNAs from cosmid and P1 clones. The PCR products were purified and sequenced using a test site protocol provided by Applied Biosystems Inc. (Foster City, CA) as previously described (Li *et al.,* 1995). Subcloned genomic DNA in Bluescript was sequenced by the dideoxy method as described previously (Li *et al.,* 1995).

FISH analysis. Metaphase chromosome preparations of lymphoblastoid cell lines were prepared by standard techniques. FISH was performed as previously described for cosmid probes (Xu *et al.,* 1994) and P1-bacteriophage probes (Albertsen *et al.,* 1994). In brief, chromosomes were first G-banded using trypsin/Wright stain, and representative cells were photographed. Slides were destained in fixative and then used for hybridizations. Genomic clones were labeled with biotin using the BioNick kit labeling system (BRL, Gaithersburg, MD). Approximately 250 ng of probe and 15 μg of human Cot-1 DNA were mixed, denatured at 75°C, and preannealed at 37°C for 1–4 h before placement on the slide. After hybridization at 37°C for 16 h, the slides were washed in 50% formamide at 37°C and 2× SSC at 47°C for both P1 and cosmid clones. Biotin signals were detected using fluorescein-conjugated avidin and amplified with biotinylated goat anti-avidin, followed by a second layer of fluorescein–avidin. Fluorescence was visualized using a Olympus BX50 epifluorescence microscope equipped with filter sets specific for FITC. Chromosomes were counterstained with propidium iodide and the metaphases that were photographed by banding were identified and photographed showing hybridization for chromosome signals. FISH signals for each metaphase were mapped to the respective chromosomal band by visual comparison of photographs.

PCR amplification of monochromosomal somatic cell hybrids. The DNA from human/rodent somatic cell hybrid cultures with a reduced number of human chromosomes (Repository No. NA 10868) was obtained from Coriell Cell Repositories (Human Genetic Mutant Cell Repositories, Camden, NJ). PCR amplification of exons 1–27b from DNA from monochromosomal somatic cell hybrids and from mouse and hamster, depending on the rodent background, was performed using primer pairs previously described (Li *et al.,* 1995). PCRs were performed on a Perkin–Elmer Cetus thermocycler. In general, all of the PCRs were performed for 30 cycles, in a final volume of 50 μl. The monochromosomal somatic cell hybrid DNA template used was 50 ng in a final volume of 50 μl except for exons 8, 9, 10a, 10c, and 23-2, where the concentration of the DNA template used was 50 ng in 25 μl. A master mix was made for 25 tubes that had the following components at final concentrations: 1× PCR buffer (Gibco–BRL, Grand Island, NY), 0.2 m*M* each dNTP, 1.5 m*M* $MgCl_2$, 0.5 μM upstream and downstream primers, and water up to 25 or 50 μl. The PCR mixes were assembled on ice using a dedicated set of pipettes, and the mixture was UV irradiated at 254 nm for 10 min (Stratalinker). The mixture was aliquoted into microfuge tubes, and 1 unit of *Taq* polymerase and the appropriate DNA template were added. The cycling conditions consisted of initial denaturation at 94°C for 5 min followed by 30 cycles of denaturation at 94°C for 1 min, 56°C for 1 min, and 72°C for 1 min followed by a final extension of 72°C for 5 min.

RESULTS

Identification and Characterization of Genomic Clones

We used *NF1* cDNA segments to isolate genomic clones from cosmid and P1-bacteriophage libraries to achieve full coverage of the *NF1* locus as a contig. Approximately 400 kb of genomic DNA is encompassed by either cosmid or P1 clones, yet, even though all 60 *NF1* exons are represented in genomic clones from this contig (Li *et al.,* 1995), a gap remains in intron 1. To fill this gap and further refine the contig map in the 5′ end of *NF1,* a PCR-based approach was used to identify the extent of exon representation in each of the genomic clones. Using intron-based primer pairs flanking each exon, we PCR-amplified genomic DNA from all of the clones in the contig. The presence of the appropriate size band as detected by agarose gel electrophoresis confirmed the presence of that specific exon and enabled us to compile the *NF1* exon content of each clone.

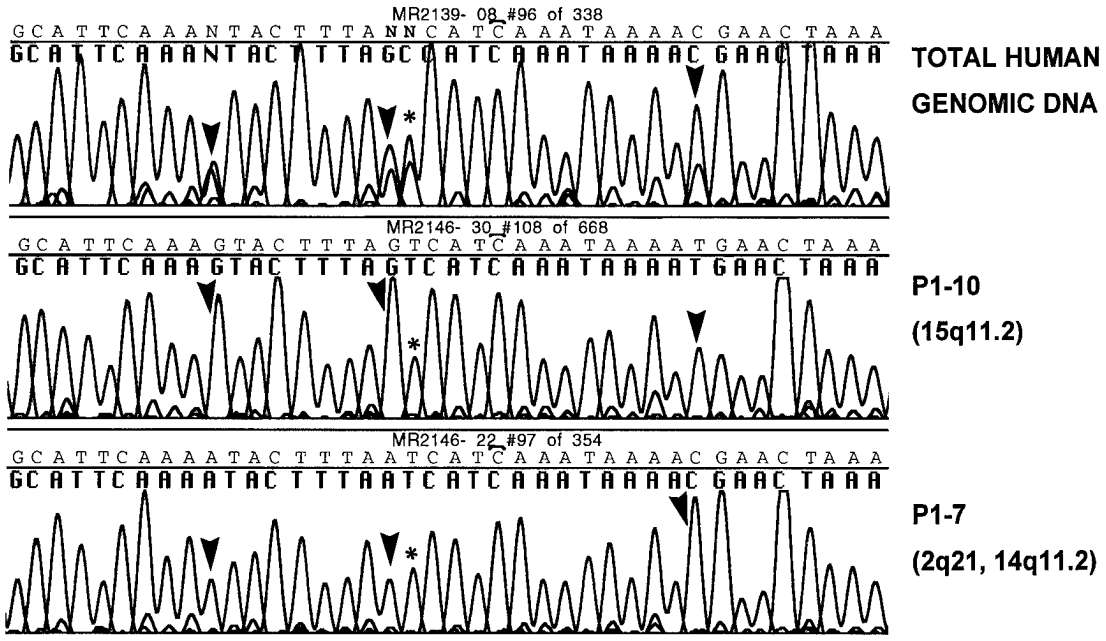


FIG. 1. An example of sequence analysis of a PCR product flanking exon 18 using total human genomic DNA and DNA from P1 clones P1-10 and P1-7 as template. Total human genomic DNA shows coamplification of more than one base at the positions marked by arrowheads. The first arrowhead shows coamplification of a G and an A and P1-10 shows a G, while P1-7 shows an A. Similarly, the second arrowhead shows coamplification of a G and an A and P1-10 shows a G, while P1-7 shows an A. The third arrowhead shows coamplification of a T and a C, and P1-10 shows a T, while P1-7 shows a C. The base marked by an asterisk shows coamplification of a T and a C, and P1-10 and P1-7 both show a C; the T comes from chromosome 22 not represented by either of these clones, but identified by PCR amplification of somatic cell hybrids.

We used the same primer pairs in *NF1*-mutation analysis to amplify uncloned total genomic DNA template, and sequences from some PCR products were ambiguous even though 4% agarose gels showed only single bands. In our efforts to optimize the sequencing protocols from uncloned genomic DNA template, we sequenced PCR products from the various genomic clones, each of which represent single alleles. Exon DNA sequence from PCR products synthesized with intron-based primers using cos 30 and P1-9 DNA template was identical to *NF1* cDNA sequence for multiple exons. However, using identical analysis, other cosmid and P1 clones, namely P1-4, P1-7, P1-15 and cos A, which were originally identified with *NF1* cDNA probes, revealed discrepancies with the *NF1* cDNA sequence. Figure 1 shows an example of sequence analysis of a PCR product flanking exon 18 using total human genomic DNA and DNA from P1 clones P1-10 and P1-7 as template. Total human genomic DNA shows coamplification of more than one base at the positions marked by arrowheads, and both bases are represented by the two P1 clones, which differ from the *NF1* locus at those sites. To characterize these separate loci better, we applied the technique of fluorescence *in situ* hybridization to map various *NF1* cDNA-selected genomic clones to their respective loci on metaphase chromosome spreads.

Fluorescence in Situ Hybridization

FISH analysis was performed for two cosmids, cos A and cos 30, and eight P1-bacteriophage clones, P1-4, P1-7, P1-9, P1-10, P1-12, P1-15, P1-16, and P1-60. Examples of FISH mapping studies are shown in Fig. 2 for three P1 clones, P1-10, P1-16, and P1-7. The left-hand photo in Fig. 2A shows strong hybridization of P1-10 to a D group chromosome, and the right-hand photo shows a G-banding pattern of the same metaphase spread, which enabled us to map the locus to chromosome 15 at band 15q11.2. Twenty cells that showed signal on both homologues were evaluated for the number of chromatids with signal (expect four per cell) to estimate the efficiency of hybridization, which was 63% for P1-10. The left-hand photo of Fig. 2B shows hybridization of P1-16 to an A group chromosome, and the right-hand photo locates the map position to bands q33–q34 on chromosome 2. The efficiency of hybridization is 53% for P1-16. The FISH photo in Fig. 2C shows weaker hybridization signal of P1-7 to a group A chromosome and a group D chromosome, and

FIG. 2. FISH analysis on the left and G-banded chromosome analysis on the right for identical metaphase spreads. (**A**) FISH mapping of P1-10 to chromosome 15q11.2. (**B**) FISH mapping of P1-16 to chromosome 2q33–q34. (**C**) FISH mapping of P1-7 to chromosomes 2q21 and 14q11.2.

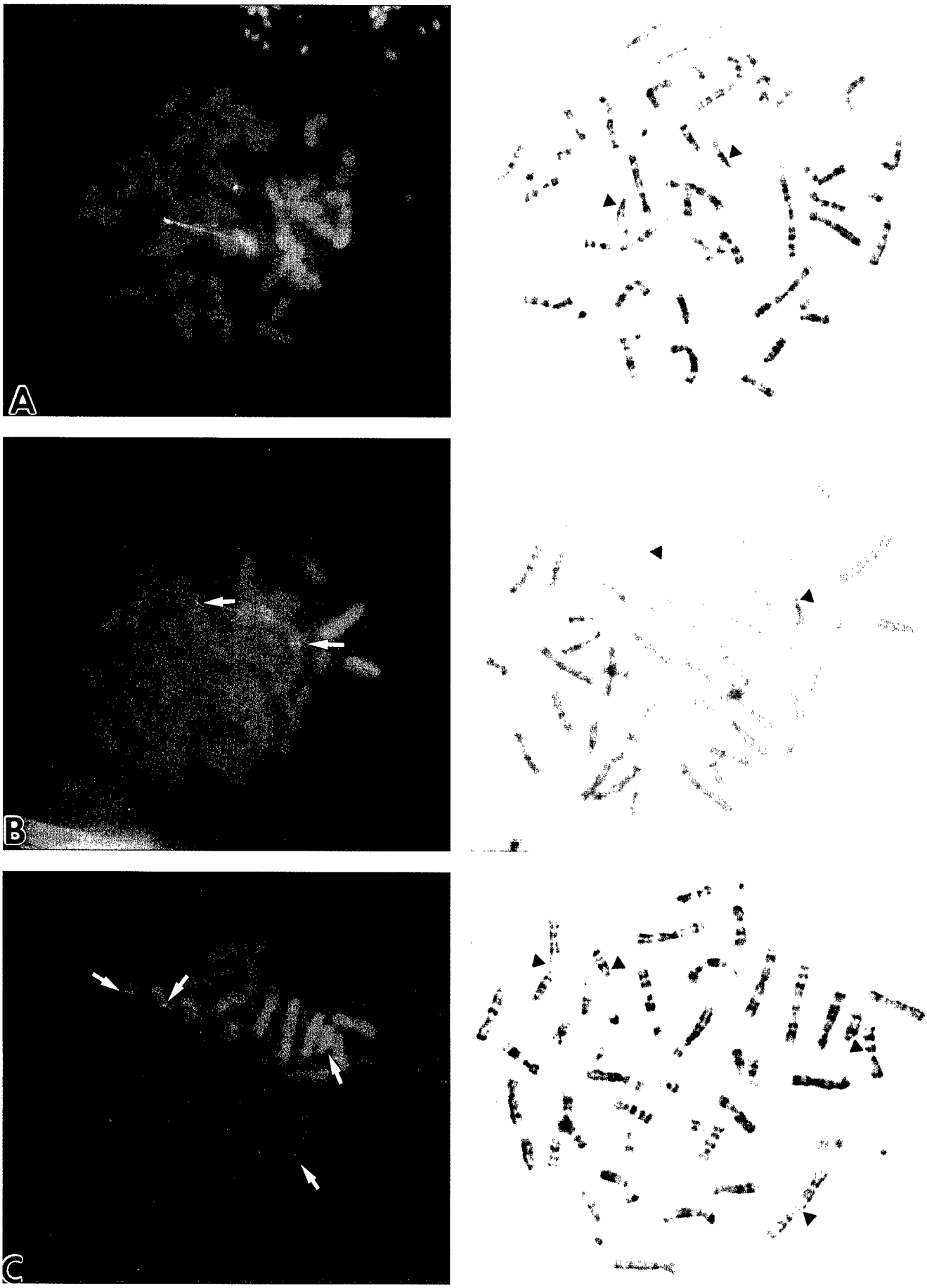

A
B
C

TABLE 1

Chromosomal Location of Genomic Clones That Were Selected by High-Stringency Hybridization Using *NF1* cDNA Segments as Probes

Genomic clone	Chromosomal location	Method of determination	*NF1* exons
Cos HB5	17q11.2	Seq, SBA	5′UTR-1
P1-60	17q11.2	FISH	5′UTR-1
Cos 30	17q11.2	Seq, FISH	2-6
P1-9	17q11.2	Seq, FISH	2-11
Cos 10	17q11.2	Seq	10c-23-1
Cos 23	17q11.2	Seq	23-2-27b
P1-12	17q11.2	FISH	Intron 27b
Cos 7	15	Seq	
Cos A	18p11.2, 21p11.2	FISH	
P1-16	2q33–q34	FISH	
P1-15	2q33–q34, 14q11.2, 22q11.2	FISH	
P1-7	2q21, 14q11.2	FISH	
P1-4	15q11.2	FISH	
P1-10	15q11.2	FISH	

Note. Seq, sequence identity by comparison; FISH, fluorescence *in situ* hybridization; SBA, Southern blot analysis. The right-hand column shows exon representation of the *NF1*-specific genomic clones.

the adjacent photo maps the FISH signals to chromosome 2 at band 2q21 and chromosome 14 at band 14q11.2. The efficiency of hybridization was estimated to be 99% for the chromosome 2 locus and 93% for the chromosome 14 locus for 20 metaphases evaluated. The chromosomal location of other clones mapped by FISH analysis is shown in Table 1. These loci have been designated as *NF1* loci (NF1-L) and they map to chromosome 2, *NF1L2* at 2q21 and *NF1L3* at 2q33–q34, chromosome 14, *NF1L4* at 14q11.2, chromosome 15, at 15q11.2, chromosome 18, *NF1L5* at 18p11.2, chromosome 21, at 21q11.2–q21, and chromosome 22, *NF1L6* at 22q11.2. The loci on chromosome 15 and 21 have been previously designated, as they have also been reported by other authors.

Monochromosomal Somatic Cell Hybrid PCR Analysis

NF1-homologous loci were also characterized by PCR amplification and direct sequencing of PCR products from monochromosomal somatic cell hybrids. Using DNA extracted from somatic cell hybrids containing single human chromosomes in a rodent background, we attempted amplification from each monochromosomal somatic cell hybrid with a combination of both intron-based and exon-based primers corresponding to *NF1* exons 1 through 27b. Exon-based primers over different segments of the *NF1* gene did not identify intronless (processed) pseudogenes. Each intron-based, exon-specific primer set was used to PCR amplify monochromosomal somatic cell hybrid DNA. The exonic segments of the *NF1* gene that coamplify from both the *NF1* locus and other chromosomes are listed in Table 2, and the respective sequence differences are also presented. In summary, PCR amplification of somatic cell hybrids confirmed the FISH mapping studies for chromosomes 2, 14, 15, 18, 21, and 22 and identified other homologous loci on chromosomes 12 and 15.

NF1-Related Loci on Chromosome 15

We used PCR amplification of somatic cell hybrids and FISH analysis to demonstrate two loci on chromosome 15. We previously identified one locus by evaluation of an SSCP variant PCR product. The primers FLR5, 5′-GTTAGCCAGCGTTTCCCTCAG-3′, and FLR3, 5′-GATTCTAGGTGGTGGCTTTTTATC-3′, PCR amplify a 129-bp product spanning a segment of exon 24 in the *NF1*-GRD that encodes the highest density of amino acid residues completely conserved in neurofibromin, p120GAP, and the yeast IRA proteins. When this primer pair was used against total human genomic DNA, a clean product of the expected size was seen on agarose gel electrophoresis. However, when the products were analyzed by single-stranded conformer analysis (PCR-SSCP, Orita *et al.,* 1989), a single com-

TABLE 2

Compilation of *NF1* Seqments That Coamplify by PCR and Show Sequence Differences between the *NF1* Locus and Homologous Loci from the Various Chromosomes

Exon	Locus	Sequence alterations
7	chr 21	9 substitutions
8	chr 18	8 substitutions
	chr 21	4 substitutions, 3-bp del
9	chr 18	3 substitutions
	chr 21	4 substitutions
11	chr 21	2 substitutions
13	chr 2	7 substitutions, 1-bp ins (A), 2-bp del
	chr 14	11 substitutions, 1-bp ins (T), 2-bp del
15	chr 14	4 substitutions
	chr 15	5 substitutions
16*	chr 12	14 substitutions, 1-bp del (G)
	chr 14	15 substitutions
17	chr 14, 22	Sequence not available
18	chr 14	9 substitutions
	chr 15	6 substitutions, 1-bp del
	chr 22	6 substitutions
19a	chr 14, 22	Sequence not available
19b	chr 15	3 substitutions, 4-bp ins
20	chr 15	11 substitutions
21	chr 15	14 substitutions, 1-bp del, 3-bp del
22	chr 15	Sequence not available
23-1	chr 15	3 substitutions
24	chr 15 (q11.2)	10 substitutions
	chr 15 (q24–qter)	8 substitutions
27a	chr 15	8 substitutions, 1-bp del
27b	chr 14	Sequence not available
	chr 15	12 substitutions, 3-bp del

Note. For exon 16 marked by an asterisk, alterations are reported in comparison with *NF1* cDNA position 2455 to 2745. Sequence from each locus has been submitted to GenBank.

plex pattern was obtained from all individuals tested. When an *NF1* genomic clone from this region, cos 10, was used as a template, a subset of the complex pattern was seen. This suggested that a locus elsewhere in the genome was coamplifying along with *NF1*. Direct sequencing of PCR products prepared from each SSC band revealed two sequences, one matching the known *NF1* sequence, the other differing by basepair substitutions at seven sites in exon 24 (bases 4111 through 4269): at bases 4122 (G to A), 4140 (C to T), 4144 (G to A), 4151 (T to C), 4179 (C to G), 4187 (C to A), and 4199 (C to T). This locus was mapped to chromosome 15 by performing the same PCR-SSCP analysis on a panel of human/rodent somatic cell hybrids from BIOS (New Haven, CT). Sublocalization within chromosome 15 was not possible using this hybrid system; however, this sequence is identical to that of NF1-HHS (Gasparini *et al.,* 1993), which was localized to q24 to qter from chromosome 15. The demonstration of an *NF1*-related locus on chromosome 15 at band q11.2 by FISH, using P1-4 and P1-10 clones as probes, predicted the presence of a second separate locus on this chromosome. The sequence of a PCR product for exon 24 from P1-4 differs by three bases from the NF1-HHS sequence at bases 4117 (C/A), 4188 (C/T), and 4244 (A/G), which supports the hypothesis that there are two loci on chromosome 15. We compared sequences from exons 20, 21, and 27a between clone P1-4 and the *NF1*-related locus, and the sequences were identical for 541 bp, which suggests that the *NF1*-related locus (Legius *et al.,* 1992) maps to 15q11.2. However, there are significant sequence differences between the *NF1*-related locus and P1-4 in exon 24, where only 6 bases are identical. This observation raises the possibility that these two loci are also distinct.

Development of Primer Pairs for NF1 Mutation Analysis

To evaluate the usefulness of our intron-based, exon-specific primer pairs in *NF1* mutation analysis, we have examined PCR amplification of uncloned genomic DNA for exons 1 through 27b, using previously published *NF1* primer sets (Li *et al.,* 1995). Some primer pairs coamplify from *NF1* and homologous loci; therefore, based on sequence differences, we redesigned most primer sets to be *NF1*-specific (Table 3).

Primer pairs for exons 8, 9, 15, 21, 27a, and 27b do not specifically amplify *NF1* from uncloned genomic DNA; therefore, to remain useful in mutation detection analysis, we attempted to separate PCR products amplified from different loci both by denaturing polyacrylamide gel electrophoresis and SSCP. Sequence analysis of aberrant bands revealed the presence of basepair substitutions, deletions, and insertions among exon and intron sequence in the homologous loci; thus, these loci likely represent nonprocessed pseudogenes. Radiolabeled PCR products amplified from total human genomic DNA, DNA specific to chromosome 17, and DNA from the homologous loci from the respective monochromosomal somatic cell hybrids were subjected to denaturing gel electrophoresis to detect size shifts. As part of this analysis the absence of amplification from the rodent background was documented for all exons except exon 23a, where we were able to separate human from amplified rodent DNA by SSCP analysis. Products for exons 21, 27a, and 27b from the individual chromosomes demonstrated fragment size variants of 1–3 bp that were easily separated on denaturing gels. For example, as shown in Fig. 3 for 27a, the PCR product from total human DNA amplified from both chromosome 17 and the homologous locus on chromosome 15. The difference in migration of the PCR product represents a basepair deletion in the chromosome 15 locus for exon 27a, which was confirmed by sequence analysis showing a basepair deletion of guanine at position 4451 in the *NF1* cDNA sequence. In addition, we performed single-stranded conformer analysis (Orita *et al.,* 1989) for the products that did not show a size shift on denaturing gel electrophoresis (exons 9 and 15). SSCP analysis of these products demonstrated the presence of aberrant bands indicative of multiple basepair substitutions, which, by DNA sequencing, showed 4 basepair substitutions in exon 9 and 5 basepair substitutions in exon 15. Using these techniques, *NF1* mutations represented by aberrant bands distinct from defined homologous loci could be identified, even when primers coamplify *NF1*-homologous loci.

DISCUSSION

Even though *NF1*-related loci had previously been reported for sequences in the 5′ region of the *NF1* gene, we were surprised by the number of non-*NF1* genomic clones that were isolated under normal stringency screening conditions. More surprising was the observation that clones from different loci amplified equivalently sized products (by 4% agarose gels) when intron-based, *NF1* exon-specific primer pairs were used in PCR. Only when PCR products were sequenced did it become clear that many clones represented *NF1*-homologous loci.

The clones with the divergent sequences mapped to different loci based on fluorescence *in situ* hybridization analysis. Given the degree of homology at multiple exons, including flanking intron sequences between numerous genomic clones representing different loci, it was surprising that the majority of clones hybridized to specific chromosomal regions by FISH without cross-hybridizing to other loci. Three clones specifically hybridized to a single locus, while three other non-*NF1* clones hybridized to multiple loci: cos A represents the homologous locus on 21q11.2 by sequence identity and cross-hybridizes to 18p11.2; P1-7, which likely represents NF1L2 at 2q21, cross-hybridizes to 14q11.2; and P1-15 hybridizes to 2q33–q34, 14q11.2, and 22q11.2.

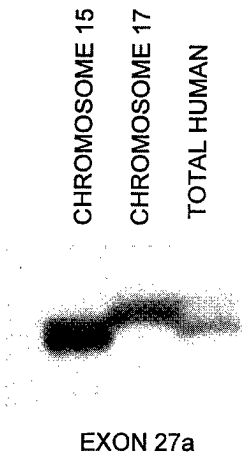


FIG. 3. Denaturing gel electrophoresis of a PCR product for exon 27a from total human genomic DNA and DNA from monochromosomal somatic cell hybrids for chromosomes 15 and 17. The product using total human genomic DNA as a template shows coamplification from the locus on chromosome 15 and chromosome 17. The product from the chromosome 15 locus shows the presence of a deletion, indicated by the shorter length of the product in comparison with the locus on chromosome 17, which on sequencing was shown to be a single base deletion of guanine (position 4451 in the *NF1* cDNA). The primer pair used is I27aUP and I27aRP (Li *et al.,* 1995). There was no amplification of rodent DNA using this primer set.

have been identified by PCR of somatic cell hybrids; neither locus is represented in our set of FISH-mapped genomic clones.

The most extensively characterized *NF1*-homologous locus is the *NF1*-related locus on chromosome 15 (Legius *et al.,* 1992). In addition, a separate locus on chromosome 15 (NF1-HHS at 15q24–qter), which differs from that described by Legius *et al.* (1992), has been reported by Gasparini *et al.* (1993) based on sequence divergence of a single PCR product amplified by *NF1* primers derived from sequences in a segment encompassing exon 24. By FISH analysis we demonstrate a chromosome 15 locus that maps to 15q11.2 using both P1-4 and P1-10 clones as probes. Comparison of the *NF1*-homologous locus on 15q11.2, represented by sequence from the two clones P1-4 and P1-10, to the *NF1*-related locus on chromosome 15 showed complete identity for exons 20, 21, and 27a, which supports the localization of the *NF1*-related locus (Legius *et al.,* 1992) to 15q11.2. However, the sequence from the *NF1*-related locus cDNA diverges remarkably from the sequence represented by P1-4 for exon 24. We hypothesize either that this may represent a third locus on chromosome 15 or that the exon 24 sequence depicted from the *NF1*-related locus may not represent the genomic sequence. The sequence reported by Legius *et al.* (1992) is a PCR product from cDNA synthesized by reverse transcription of total RNA isolated from cultured human/rodent somatic cell hybrid cells containing chromosome 15 as the human complement. It is possible that the primary transcript from this homologous locus is processed differently and exon 24 is aberrantly spliced. Furthermore, the *NF1*-related locus sequence contains only 12 bases of exon 24, and only 9 of those 12 are identical with *bona fide NF1* cDNA sequence. This raises the possibility that this sequence may not truly represent exon 24 sequence from the *NF1*-related locus at the genomic level. We have also independently confirmed the existence of homologous loci on chromosomes 14 and 22 (Marchuk *et al.,* 1991) and chromosome 21 at 21q11.2 (Suzuki *et al.,* 1994) by FISH and PCR of somatic cell hybrids.

Genetic variation within pseudogenes may be important reservoirs that cause mutations by homologous recombination between pseudogenes and genomic segments encoding functional domains. Conversions between gene and pseudogene have been previously reported for a number of genes (Braun *et al.,* 1990; Sorge *et al.,* 1990; Urabe *et al.,* 1990; Matsuno *et al.,* 1992; Collier *et al.,* 1993; He and Grabowski, 1993), all of which represent linked loci. The only example of gene conversion for unlinked loci was reported for the von Willebrand factor gene on chromosome 12 and its pseudogene on chromosome 22 (Eikenboom *et al.,* 1994). This raises the hypothesis that homologous recombination involving intragenic segments between *NF1* and *NF1*-homologous loci could be the etiology of some *NF1*-disease causing mutations (Cummings *et al.,* 1993). Mutation by gene conversion predicts that some *NF1* mutations would correspond to sequences from *NF1*-homologous loci. We have examined 21 known *NF1* mutations (NNFF Mutation Analysis Consortium Newsletter, June 1995) that lie in exons for which we have detected *NF1*-homologous segments. Only one mutation, R1513X in exon 27a, corresponds to a base change in an *NF1*-homologous locus on 15q11.2; however, this mutation occurs at a highly mutable CpG site. Thus, this limited analysis provides no direct evidence of *NF1* gene conversion.

NF1-homologous loci complicate mutation analysis of the *NF1* gene. We have detected regions homologous to exons 7, 8, 9, 11, 13, 15, 16, 17, 18, 19a, 19b, 20, 21, 22, 23-1, 24, 27a, and 27b on chromosomes 2, 12, 14, 15, 18, 21, and 22. To design *NF1*-specific primers, we sequenced intronic segments from PCR products from both genomic clones and monochromosomal somatic cell hybrids representing *NF1*-related loci and compared them with *bona fide NF1* sequence. New primers were designed to have natural mismatches with the homologous sequences at the 3′ end of the oligonucleotides. When only a single naturally occurring mismatch was available, we created an additional mismatch to increase primer specificity, whereby multiple mismatches would lead to preferential hybridization of the primer to the *NF1* locus. Thus, we have compiled a set of oligonucleotide primer pairs that amplify uncloned human DNA template specifically from the *NF1* locus (Table 3); only primer pairs for exons 8, 9, 15, 21, 27a, and 27b are nonspecific. PCR amplification using these *NF1*-primer pairs has simplified the DNA-based

NF1 allele from the bone marrow of children with type 1 neurofibromatosis and malignant myeloid disorders. *N. Engl. J. Med.* **330:** 597–601.

Sorge, J., Gross, E., West, C., and Beutler, E. (1990). High level transcription of the glucocerebrosidase gene in normal subjects and patients with Gaucher disease. *J. Clin. Invest.* **86:** 1137–1141.

Suzuki, H., Ozawa, N., Taga, C., Kano, T., Hattori, M., and Sakkai, Y. (1994). Genomic analysis of a *NF1*-related pseudogene on human chromosome 21. *Gene* **147:** 277–280.

The, I., Murthy, A. E., Hannigan, G. E., Jacoby, L. B., Menon, A. G., Gusella, J. F., and Bernards, A. (1993). Neurofibromatosis type 1 gene mutations in neuroblastoma. *Nature Genet.* **3:** 362–366.

Urabe, K., Kimura, A., Harada, F., Iwanga, T., and Sasazuki, T. (1990). Gene conversion in 21-hydroxylase genes. *Am. J. Hum. Genet.* **46:** 1178–1186.

Viskochil, D., Buchberg, A. M., Xu, G., Cawthon, R. M., Stevens, J., Wolff, R. K., Culver, M., Carey, J. C., Copeland, N. G., Jenkins, N. A., White, R., and O'Connell, P. (1990). Deletions and translocations interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell* **62:** 187–192.

Wallace, M. R., Marchuk, D. A., Andersen, L. B., Letcher, R., Odeh, H. M., Saulino, A. M., Fountain, J. W., Breeton, A., Nicholson, J., Mitchell, A. L., Brownstein, B. H., and Collins, F. S. (1990). Type 1 neurofibromatosis gene: Identification of a large transcript disrupted in 3 NF-1 patients. *Science* **249:** 181–186.

Xu, G., O'Connell, P., Viskochil, D., Cawthon, R., Robertson, M., Culver, M., Dunn, D., Stevens, J., Gesteland, R., White, R., and Weiss, R. (1990a). The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell* **62:** 599–608.

Xu, G., Lin, B., Tanaka, K., Dunn, D., Wood, D., Gesteland, R., White, R., Weiss, R., and Tamanoi, F. (1990b). The catalytic domain of the neurofibromatosis type I gene product stimulates rasGTPase and complements ira mutants of S. cerevisae. *Cell* **63:** 835–841.

Xu, P., Zhu, X. L., Hueckstedt, T. P., Brothman, A. R., and Hoidal, J. R. (1994). Assignment of human xanthine dehydrogenase gene to chromosome 2p22. *Genomics* **23:** 289–291.

Xu, W., Mulligan, L. M., Ponder, M. A., Liu, L., Smith, B. A., Mathew, C. P. G., and Ponder, B. A. J. (1992). Loss of NF-1 alleles in phaeochromocytomas from patients with type-1 neurofibromatosis. *Genes Chrom. Cancer* **4:** 337–342.

GENOME METHODS

Long RT–PCR of the Entire 8.5-kb *NF1* Open Reading Frame and Mutation Detection on Agarose Gels

Jennifer M. Martinez,[1] Heidi Huntsman Breidenbach,[1] and Richard Cawthon[1]

[1]Department of Human Genetics, [2]Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112

6:58–66 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

(ORF) and identifed mutations in 14 of 21 NF1 patients (67%). A similar detection rate in NF1 patients was obtained with the PTT in our own laboratory (H. Breidenbach, unpubl.)

Because of limitations in translating large proteins in vitro and in resolving large proteins on polyacrylamide gels, PTT currently has an upper limit of ~2 kb for the largest PCR product that can be screened effectively. Therefore, the 8.5-kb ORF of *NF1* must be divided into overlapping segments for amplification. This limitation means that deletion mutations large enough to span the entire region of overlap of two consecutive cDNA segments that are the targets for amplification will prevent amplification from the mutant allele, and these mutations will be missed.

Recently, modified reaction conditions that allow PCR amplification of much longer DNA targets were described (Barnes 1994; Cheng et al. 1994). Long RT–PCR of the entire *NF1* ORF followed by agarose gel electrophoresis, presented here, allows identification of a great variety of large deletions that would be missed in smaller, overlapping RT–PCR products. Furthermore, small insertion/deletion mutations present in many of the patients that show apparently normal long RT–PCR products can be detected conveniently by digestion of the long RT–PCR product with *Ase*I and *Fsp*I, followed by agarose gel electrophoresis.

RESULTS

Figure 1 summarizes our strategy for mutation detection in long RT–PCR products from *NF1*. Total RNA from Epstein–Barr virus (EBV)-transformed lymphoblasts or from whole blood was the template for RT–PCR. Previous work has shown that these cells express both the (+) 23A and the (–) 23A forms of the *NF1* transcript, but do not express either (+) 9A or (+) 48A transcripts. Therefore, the long RT–PCR method presented here should amplify, from control individuals, only *NF1* (–) 9A (+) 23A (–) 48A (8754 bp) and *NF1* (–) 9A (–) 23A (–) 48A (8691 bp).

Reverse transcription is primed with a specific antisense oligonucleotide downstream of the *NF1* stop codon. PCR from the resulting cDNA is then performed using a pair of primers flanking the ORF. The product of this long RT–PCR is screened for deletions and insertions by agarose gel electrophoresis. To allow detection of smaller deletions and insertions and to narrow the location of all detected mutations, the long product is then digested with restriction endonucleases that release a set of fragments with unique sizes that are well separated by agarose gel electrophoresis. The restriction fragments predicted for an *Ase*I–*Fsp*I double digest (using the Site Program from the Intelligenetics Suite) appeared suitable for this purpose and are shown in Figure 1.

To test its effectiveness, we applied this approach to mutation detection to all of the patients in our NF1 collection with known deletions or insertions 30 bp or longer that fall entirely between exons 1 and 49 of the *NF1* gene. The mutations in *NF1* mRNA of the nine patients constituting this set have not been reported previously; their sizes and locations are shown in Figure 1 and Table 1. Table 1 also lists the observed and expected sizes of the mutant bands before and after restriction digestion, and the normal restriction fragments affected by the mutations.

Figure 2 shows the results of gel electrophoresis in 0.8% agarose of long *NF1* RT–PCR products prepared from the nine NF1 patients and three controls. A band of ~8.7 kb was shared by all samples. Because the (+) and (–) 23A forms of *NF1* differ in length by only 63 bp, resolution of the mixed RT–PCR products into two bands (8754 and 8691 bp, respectively) is not expected under these conditions. Agarose gel electrophoresis of long *NF1* RT–PCR product prepared from whole blood RNA of a control gave results equivalent to those shown in the control lanes in Figure 2.

Figure 3 shows the patient samples and a control after double digestion with *Ase*I and *Fsp*I and additional electrophoresis in a 1.5% agarose gel until the bromophenol blue had migrated nearly to the bottom of the gel (19 cm). The six restriction fragments expected from normal *NF1* RT–PCR products were observed in all samples. The largest two fragments, Ia and Ib, appeared to have a size difference consistent with alternative splicing of the 63-bp exon 23A. [All restriction digest data presented here were obtained using *Fsp*I and *Ase*I. However, we now recommend that *Fsp*I be replaced by the isoschizomer *Avi*II (see Methods)].

One of the difficulties of mutation detection in RNA samples from NF1 individuals arises from the fact that 80% of patients show unequal levels of mRNA from the two *NF1* alleles (Hoffmeyer et al. 1995). In two-thirds of the patients tested, the abundance of mRNA from one *NF1* allele was more than twice that of the other allele. That study used heterozygosity at a single-base-pair

Table 1. NF1 Patients, Their Known Mutations, and Mutant Band Sizes Determined by Agarose Gel Electrophoresis Before and After Restriction Digestion with *AseI–FspI*

Lane no.	Patient no.	Mutation	Size of mutant band by AGE: observed (predicted)[a]	Restriction fragment affected	Size of mutant band by AGE after restriction digest: observed (predicted)[a]
1	11387	del exon 37 (102 bp)	—	II	2189 (2182)
2	119-01	del 30 bp from exon 9	—	IV	918 (921)
3	11389	ins 176 bp into exon 9	—	IV	1126 (1127)
4	12359	del exon 35 (62 bp)	—	II	2229 (2222)
5	11573	del exons 28–31 (1171 bp)	7578 (7583 + 23A; 7520, −23A)	Ia, Ib	1840 (1836, + 23A; 1773, −23A)
6	11360	del exons 28–39 (2354 bp)	6390 (6400, +23A; 6337, −23A)	Ia, Ib, II	2880 (2937, +23A; 2874, −23A)
7	11344	del exon 3 (84 bp)	—	V	627 (628)
8	107-06	del exon 20 (182 bp)	—	Ia, Ib, III	4604 (4625, +23A; 4562, −23A)
9	94-04	del 148 bp from exon 29	—	Ia	2865 (2859, +23A)
				Ib	2802 (2796, −23A)
9	94-04	del exon 29 (341 bp)	8439 (8413, +23A; 8350, −23A)	Ia	2682 (2666, +23A)
				Ib	2624 (2603, −23A)

[a]Sizes of the observed mutant bands are given. The numbers in parentheses to the right of the observed sizes are the sizes predicted based on previous independent analyses of the mutations. (AGE) Agarose gel electrophoresis. Band sizes were determined with the FragmeNT Analysis Program (Molecular Dynamics, Inc.).

the mutant band (Table 1) from the average size for the normal (±) 23A RT–PCR products (8723 bp). The approximate location of these deletions within the long RT–PCR product can be deduced from the sizes of the abnormal fragments generated by restriction enzyme digestion (Fig. 3). In patient 11573, deletion of a single 1.1-kb segment from the *AseI–FspI* restriction map in Figure 1 could produce the 1.8-kb mutant fragment observed after restriction digestion (Fig. 3, lane 5) only if the deletion falls within fragment I. Similarly, in patient 11360, a 2.3-kb deletion could produce the 2.9-kb mutant fragment observed after restriction digestion (Fig. 3, lane 6) only if the deletion takes out the *AseI* site at the fragment I/fragment II boundary.

Previous work has shown that the mutation in patient 11573 is a deletion of 11 kb of genomic DNA removing exons 28–31, and the mutation in patient 11360 is a deletion of 40 kb of genomic DNA removing exons 28–39 (Viskochil et al. 1990; R. Weiss, GenBank accession no. LO3723). In both cases, because the deleted regions are well away from the region of exon 23A, both the (+) and (−) 23A forms of *NF1* mRNA are expected to be synthesized from the mutant alleles. Because of the small size of the length differences, resolution of the (+) and (−) 23A mutant bands by gel electrophoresis prior to restriction digestion was not expected. After restriction digestion, however, two mutant fragments 63 bp apart were expected from each of these two patients. Only a single mutant restriction fragment from each patient was observed (Fig. 3). It is likely that the lower mutant band from patient 11573 did not resolve from the normal 1800-bp band and that the higher mutant band from patient 11360 did not resolve from the normal 3007-bp band.

Patient 94–04

Continued agarose gel electrophoresis of the initial, uncut long RT–PCR products until the full-length control products had migrated 8.5 cm from the origin revealed one additional abnor-

tion. Insertion/deletion mutations undetected by agarose gel electrophoresis of the uncut long *NF1* RT–PCR product should be <300 bp. Therefore, in these patients the mutant fragments resulting from the *Ase*I–*Fsp*I digestion usually will migrate in proximity to the corresponding normal restriction fragment, allowing easy approximate localization within the *NF1* ORF. This prediction was upheld for five of the six patients (Fig. 3; Table 1).

Exceptions to this rule include any small mutation that removes one of the restriction sites targeted for *Ase*I–*Fsp*I digestion, thereby generating a disproportionately large mutant restriction fragment. The approximate location of such mutations often can be deduced from the size of the resulting fragments. For example, a deletion of <300 bp from the *NF1* long RT–PCR product will result in a mutant fragment ~4.6 kb after restriction digestion (Fig. 3, lane 8) only if it removes the *Fsp*I restriction site in exon 20. Sequencing of small RT–PCR products spanning the exon 20 region of this patient, 107-6, has demonstrated that the mutant *NF1* mRNA is missing exactly the sequence of exon 20.

The approximate location of a mutation also can be deduced when the loss of signal intensity from the position of one or more of the normal restriction fragments, relative to the signal intensity of the other normal fragments in the same lane, is large enough to measure. The signal decrease occurs because the mutant fragment shifts in the gel lane away from the position(s) of the normal fragment(s). Examples of this are the mutations in patients 11360 and 11344. With the unaided eye, it is clear in Figure 3 that fragment II (lane 6) and fragment V (lane 7) are diminished relative to the other normal-sized fragments in the same lanes. Less obvious signal losses from normal fragments, resulting from other mutations, were identifiable on careful quantitation with Molecular Dynamics ImageQuaNT software. However, this approach will not succeed in all cases, because of the very low abundance of some patients' mutant *NF1* mRNAs.

DISCUSSION

Long RT–PCR of the entire *NF1* ORF and analysis of the products on agarose gels before and after restriction digestion with *Ase*I and *Fsp*I is shown here to be a highly effective way to identify a great variety of *NF1* mutations. In the assay reported here, for all nine patient samples analyzed, the mutant bands observed on agarose gels before and after restriction digestion were the sizes predicted based on the patients' known insertion/deletion mutations. The greatest strength of the assay is its power to yield mutant RT–PCR products from deletion mutations that would be missed by traditional methods. For example, both of the deletions in patients 11573 and 11360 would not be detected by the protein truncation assay described by Heim et al. (1995), because the binding sites are removed for both the downstream primer of the S3 segment and the upstream primer of the S4 segment.

The power of gel electrophoresis to resolve bands of similar size depends in part on the location of the bands relative to the electrophoretic front during electrophoresis (R_f = distance of band from the origin/distance of the electrophoretic front from the origin); the best resolution is an R_f near 0.5. Because the 2% difference in length between the normal fragments Ia and Ib was resolvable at an R_f of ~0.31, it is likely that the smallest possible single exon dropout from lymphoblast *NF1* mRNA, the 47-bp exon 47, would be detected also: It contributes 2.1% of the length of fragment II, which has an R_f of ~0.36 in these 1.5% agarose gels.

These considerations also suggest that a few samples with no apparent abnormalities by a single gel analysis of the *Ase*I–*Fsp*I digestion products may reveal small deletions or insertions upon re-electrophoresis of the samples on two additional gels, one with a lower agarose concentration to optimize resolution of fragments Ia, Ib, II, and III, and the other with a higher agarose concentration to optimize resolution for of IV and V.

Occasionally, mutant restriction fragments will be missed when they are of low intensity and indistinguishable in size from one of the faint background bands appearing in all samples. Because the sizes of several of these faint bands are determined by the positions of the recognition sites for the restriction enzymes, this problem may be solved by use of a second set of restriction enzymes to cut the long *NF1* RT–PCR product, in place of the *Ase*I–*Fsp*I combination.

The assay presented here should detect ~30% of the mutations reported to date to the NF1 Mutation Consortium (Korf 1995). Currently, we are in the process of testing its detection rate in a set of unrelated NF1 patients. Expressed as mutations identified per *NF1* RT–PCR product, this assay should far surpass all assays reported previously. It will be particularly useful in the detec-

Restriction Enzyme Digest and Analysis

For each RT–PCR product, the PCR buffer was exchanged with water using a Centricon-100 concentrator. Five microliters of the 8.7-kb PCR product and λ *Hin*dIII mass standards was analyzed on 1 × Tris-acetate–EDTA (TAE) 0.8% agarose gels containing ethidium bromide and photographed with transillumination by UV light. Between 200 and 870 ng (~20 μl) of each sample was then digested with 10–20 units *Fsp*I (New England Biolabs) under mineral oil in a final aqueous volume of 50 μl containing 50 mM Tris-HCl (pH 8.0) and 10 mM $MgCl_2$, at 37°C for 4 hr. The enzyme was then heat inactivated at 68°C for 20 min. Between 10 and 20 units of *Ase*I (New England Biolabs) and additional reagents were then added to a final volume of 70 μl containing the following: 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 100 mM NaCl, and 1 mM DTT. Incubation was then carried out at 37°C for 4 hr, followed by heat inactivation at 68°C for 20 min. After consecutive digestion with *Fsp*I and *Ase*I, the sample was concentrated by Centricon-100 centrifugation and/or evaporative centrifugation in a Savant Speed-Vac and analyzed on a 20-cm 1 × TAE 1.5% agarose gel until the bromophenol dye front had traveled nearly the length of the gel. Following electrophoresis, the gel was stained with a 1:10,000 dilution of stock Sybr Green I nucleic acid gel stain (Molecular Probes, Inc.) in 1 × TAE for 15 min and scanned using a FluorImager (Molecular Dynamics) with the 530 DF30 yellow filter in place.

The data in Figure 3 were obtained using the procedure described above for double digestion of the *NF1* RT–PCR product. Recently, however, we were successful in greatly simplifying the protocol for restriction digestion. *Avi*II (Boehringer-Mannheim) is an isoschizomer of *Fsp*I that has a recommended buffer for digestion that is identical to that of *Ase*I (New England Biolabs). We therefore replaced *Fsp*I with *Avi*II, and found that we could digest the *NF1* long RT–PCR product simultaneously with a mixture of *Avi*II and *Ase*I. In our current protocol, upon completing the long PCR step, 100–200 ng of PCR product (10–20 μl) was diluted directly into a final volume of 25–30 μl containing 1 × *Ase*I buffer (New England Biolabs), 4 mM spermidine, 1 mM DTT, and 10 units each of *Avi*II and *Ase*I. Incubation was done at 37°C for 1 hr. The digestion products were then analyzed by agarose gel electrophoresis as described above.

Estimation of Band Sizes on Agarose Gels

All length determinations of gel bands were performed using the Molecular Dynamics Fragment Analysis software, version 1.1, running under Microsoft Windows NT on an Intel 486-based computer. The point-to-point logarithmic option for size analysis was applied. The size standard was the GIBCO–BRL 1-kb ladder.

ACKNOWLEDGMENTS

This work was supported by grant MGN 1 R55 CA57511-01 from the National Cancer Institute and grant DAMD17-93-J-3044 from the Department of the Army. We thank Roche Molecular Systems for providing us with an XL PCR kit in the early stages of this project. We thank Dr. Ray White and Dr. Ken Ward for providing lymphoblast cell lines prepared from NF1 patients; Mr. Ed Meenen for oligonucleotide syntheses; Ms. Margaret Robertson for sequencing PCR products; and Dr. David Viskochil, Dr. Smita Purandare, Dr. Shun Sawada, and Ms. Shannon Neil for helpful discussions.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Andersen, L.B., J.W. Fountain, D.H. Gutmann, S.A. Tarle, T.W. Glover, N.C. Dracopoli, D.E. Housman, and F.S. Collins. 1993. Mutations in the neurofibromatosis 1 gene in sporadic malignant melanoma cell lines. *Nature Genet.* **3:** 118–121.

Barnes, W.M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci.* **91:** 2216–2220.

Blatt, J., R. Jaffe, M. Deutsch, and J.C. Adkins. 1986. Neurofibromatosis and childhood tumors. *Cancer* **57:** 1225–1229.

Cawthon, R.M., R. Weiss, G. Xu, D. Viskochil, M. Culver, J. Stevens, M. Robertson, D. Dunn, R. Gesteland, P. O'Connell, and R. White. 1990. A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell* **62:** 193–201.

Cheng, S., C. Fockler, W.M. Barnes, and R. Higuchi. 1994. Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci.* **91:** 5695–5699.

Danglot, G., V. Regnier, D. Fauvet, G. Vassal, M. Kujas, and A. Bernheim. 1995. Neurofibromatosis 1 (NF1) mRNAs expressed in the central nervous system are differentially spliced in the 5′ part of the gene. *Hum. Mol. Genet.* **4:** 915–920.

Heim, R.A., L.N. Kam-Morgan, C.G. Binnie, D.D. Corns, M.C. Cayouette, R.A. Farber, A.S. Aylsworth, L.M. Silverman, and M.C. Luce. 1995. Distribution of 13 truncating mutations in the neurofibromatosis 1 gene. *Hum. Mol. Genet.* **4:** 975–981.

Hoffmeyer, S., G. Assum, J. Griesser, D. Kaufmann, P. Nurnberg, and W. Krone. 1995. On unequal allelic expression of the neurofibromin gene in neurofibromatosis type 1. *Hum. Mol. Genet.* **4:** 1267–1272.

Korf, B.R. 1995. *NF1 genetic analysis consortium newsletter.* KORF@A1.TCH.HARVARD.EDU.

Li, Y., G. Bollag, R. Clark, J. Stevens, L. Conroy, D. Fults, K. Ward, E. Friedman, W. Samowitz, M. Robertson, P. Bradley, F. McCormick, R. White, and R. Cawthon. 1992.